



**DISERTASI – SS143506**

**BAYESIAN *MIXTURE* MODEL AVERAGING  
UNTUK MENGIDENTIFIKASI PERBEDAAN  
EKSPRESI GEN PERCOBAAN *MICROARRAY***

**Ani Budi Astuti  
Nrp. 1311301002**

**PROMOTOR/CO-PROMOTOR:**

**Prof. Drs. Nur Iriawan, M.IKom., Ph.D  
Dr. Irhamah, S.Si, M.Si  
Dr. rer. pol. Heri Kuswanto, S.Si., M.Si**

**PROGRAM DOKTOR  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017**



## LEMBAR PENGESAHAN DISERTASI

Disertasi disusun untuk memenuhi salah satu syarat memperoleh  
gelar Doktor (Dr.)

di Program Doktor Departemen Statistika  
Institut Teknologi Sepuluh Nopember

Oleh:  
**Ani Budi Astuti**  
**NRP. 1311301002**

Tanggal Ujian : 26 Juli 2017  
Periode Wisuda: September 2017

### Disetujui oleh:

- |   |       |
|---|-------|
| 1. Prof. Drs. Nur Iriawan, M.IKom., Ph.D (Promotor)       | ----- |
| NIP. 19621015 198803 1 002                                |       |
| 2. Dr. Irhamah, S.Si, M.Si (Co-Promotor)                  | ----- |
| NIP. 19780406 200112 2 002                                |       |
| 3. Dr. rer. pol. Heri Kuswanto, S.Si., M.Si (Co-Promotor) | ----- |
| NIP. 19820326 200312 1 004                                |       |
| 4. Dr. Danardono, M.PH (Penguji Eksternal)                | ----- |
| NIP. 19680919 199412 1 001                                |       |
| 5. Dr. Santi Wulan Purnami, S.Si, M.Si (Penguji)          | ----- |
| NIP. 19720923 199803 2 001                                |       |
| 6. Dr. I Nyoman Latra, MS (Penguji)                       | ----- |
| NIP. 19511130 197901 1 001                                |       |
| 7. Dr. R. Mohamad Atok, M.Si (Penguji)                    | ----- |
| NIP. 19710915 199702 1 001                                |       |

Dekan Fakultas MIPA

Prof. Dr. Basuki Widodo, M.Sc.  
NIP. 19650605 198903 1 002

## LEMBAR PENGESAHAN DISERTASI

Disertasi disusun untuk memenuhi salah satu syarat memperoleh  
gelar Doktor (Dr.)


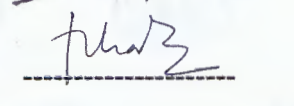

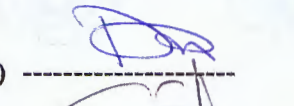
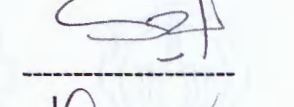
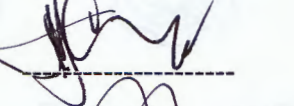
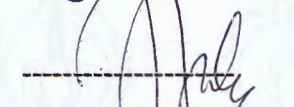
di Program Doktor Departemen Statistika  
Institut Teknologi Sepuluh Nopember

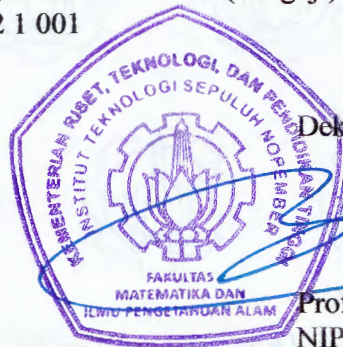
Oleh:  
**Ani Budi Astuti**  
NRP. 1311301002

Tanggal Ujian : 26 Juli 2017  
Periode Wisuda: September 2017

Disetujui oleh:

1. Prof. Drs. Nur Iriawan, M.IKom., Ph.D (Promotor)  
NIP. 19621015 198803 1 002
2. Dr. Irhamah, S.Si, M.Si (Co-Promotor)  
NIP. 19780406 200112 2 002
3. Dr. rer. pol. Heri Kuswanto, S.Si., M.Si (Co-Promotor)  
NIP. 19820326 200312 1 004
4. Dr. Danardono, M.PH (Penguji Eksternal)  
NIP. 19680919 199412 1 001
5. Dr. Santi Wulan Purnami, S.Si, M.Si (Penguji)  
NIP. 19720923 199803 2 001
6. Dr. I Nyoman Latra, MS (Penguji)  
NIP. 19511130 197901 1 001
7. Dr. R. Mohamad Atok, M.Si (Penguji)  
NIP. 19710915 199702 1 001

  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_



Dekan Fakultas MIPA

  
Prof. Dr. Basuki Widodo, M.Sc.  
NIP. 19650605 198903 1 002



## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam disertasi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Doktor di suatu Perguruan Tinggi manapun dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Surabaya, 8 Agustus 2017  
Yang menyatakan,

Ani Budi Astuti



**BAYESIAN MIXTURE MODEL AVERAGING  
UNTUK MENGIDENTIFIKASI PERBEDAAN EKSPRESI GEN  
PERCOBAAN MICROARRAY**

**Nama Mahasiswa : Ani Budi Astuti**  
**NRP : 1311301002**  
**Promotor : Prof. Drs. Nur Iriawan, M.IKom., Ph.D**  
**Co-promotor : Dr. Irhamah, S.Si, M.Si**  
**Dr. rer. pol. Heri Kuswanto, S.Si., M.Si**

**ABSTRAK**

Data ekspresi gen merupakan data hasil percobaan *microarray*. Tantangan utama analisis statistika untuk data *microarray* adalah jumlah sampel yang tersedia sedikit tetapi karakteristik variabel yang terukur sangat banyak. Hal ini disebabkan oleh keterbatasan dana, tenaga dan waktu serta ketersediaan sampel. Selain itu, sifat dasar dari sebaran data ekspresi gen yang diperoleh akan sangat kompleks (multimodal). Penelitian yang terkait dengan identifikasi perbedaan ekspresi gen telah banyak dilakukan dengan menggunakan pendekatan Bayesian. Analisis Bayesian yang digunakan adalah *Bayesian Mixture Model* (BMM) karena indikasi data memiliki distribusi *mixture*. Selain itu, secara terpisah, *Bayesian Model Averaging* (BMA) digunakan untuk mengidentifikasi perbedaan ekspresi gen dengan prinsip memprediksi model terbaik melalui rata-rata terboboti dari semua kemungkinan model baik dari data. Tujuan dari penelitian ini adalah menggabungkan metode BMM dan metode BMA sehingga terbentuk model *Bayesian Mixture Model Averaging* (BMMA) untuk data *Chickpea* di Indonesia. Hasil penelitian menunjukkan bahwa model BMMA-normal mampu dengan baik dalam mengidentifikasi kelompok ID gen data *Chickpea* dalam kelompok *Up-regulated*, *Regulated*, dan *Down-regulated* dengan tetap mempertahankan konsep data *driven*. Model BMMA-normal telah berhasil mengidentifikasi lima ID gen yang masuk dalam kelompok *Up-regulated* dan 14 ID gen masuk dalam kelompok *Regulated*. Dalam penelitian ini tidak ditemukan ID gen yang masuk dalam kelompok *Down-regulated*. Model BMMA-normal untuk kasus data *Chickpea* adalah model terbaik dibandingkan model pendekatan *non- mixture* normal. Nilai *goodness of fit* Kolmogorov-Smirnov yang diperoleh cukup kecil dan masih dalam jaminan lebih kecil dari nilai Kolmogorov-Smirnov tabel.

**Katakunci:** Bayesian, Model *Mixture*, Model *Averaging*, Model *Mixture Averaging*, *Microarray*-Indonesia.





# **BAYESIAN MIXTURE MODEL AVERAGING TO IDENTIFY DIFFERENCES GENE EXPRESSION OF MICROARRAY EXPERIMENT**

**Student Name** : Ani Budi Astuti  
**Register Number** : 1311301002  
**Supervisor** : Prof. Drs. Nur Iriawan, M.IKom., Ph.D  
**Co-supervisor** : Dr. Irhamah, S.Si, M.Si  
Dr. rer. pol. Heri Kuswanto, S.Si., M.Si

## **ABSTRACT**

Gene expression data is the data from the results of microarray experiment. The main challenge of the statistical analysis for microarray data is the small number of samples available but the characteristics of the variables measured are numerous. This is due to the limited funds, personnel, time and availability of samples. Moreover, the nature of the distribution of gene expression data obtained will be very complex (multimodal). Research related to the identification of gene expression differences have been carried out by using a Bayesian approach. Bayesian analysis is used Bayesian Mixture Model (BMM) because an indication of the data has a distribution mixture. In addition, separately, Bayesian Model Averaging (BMA) is used to identify gene expression differences with the principle of predicting the best model through the weighted average of all possible good models of data. The results showed that the BMMA-normal model able to properly identify groups of gene ID of the *Chickpea* data in the of Up-regulated, Regulated, and Down-regulated groups while maintaining the concept of data driven. BMMA-normal model have successfully identified five genes ID in the Up-regulated group and 14 genes ID in the Regulated group. In these studies could not find the genes ID in the Down-regulated group. The BMMA-normal model for Chickpea data cases is the best model compared to the normal non-mixture approach model. The value of goodness of fit Kolmogorov-Smirnov obtained is quite small and still under warranty smaller than the value of Kolmogorov-Smirnov table.

**Keywords:** Bayesian, Mixture Model, Averaging Model, Mixture Averaging Model, Microarray-Indonesia.



## KATA PENGANTAR

Alhamdulillahirabbilalamin, penulis panjatkan puji dan syukur kehadiran Allah SWT atas segala limpahan karunia, rahmat, nikmat, taufik dan hidayah-Nya sehingga penulis sanggup dan mampu menyelesaikan penyusunan Disertasi dengan judul **“Bayesian Mixture Model Averaging untuk Mengidentifikasi Perbedaan Ekspresi Gen Percobaan Microarray”**. Disertasi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Program Doktor di Jurusan Statistika FMIPA, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.

Keberhasilan dalam menyelesaikan penyusunan disertasi ini tidak terlepas dari semangat, bimbingan, dorongan, kerjasama, bantuan dan motivasi dari berbagai pihak, baik langsung maupun tidak langsung. Oleh karena itu pada kesempatan ini, sebagai rasa syukur dan ucapan terimakasih serta penghargaan yang tinggi penulis sampaikan kepada:

1. Rektor UB, Dekan FMIPA UB, Ketua Jurusan Matematika UB dan Ketua Program Studi Statistika UB yang telah memberikan kesempatan dan fasilitas kepada penulis untuk melanjutkan studi ke jenjang pendidikan Program Doktor Statistika di ITS Surabaya.
2. Direktorat Jenderal Pendidikan Tinggi, Kementerian Ristek dan Teknologi Nasional yang telah memberikan bantuan dana beasiswa BPPS untuk S3.
3. Dr. Suhartono, M.Sc, sebagai Ketua Jurusan Statistika ITS yang sekaligus sebagai Dosen Penasehat Akademik dan Dr. rer. pol. Heri Kuswanto, S.Si., M.Si, sebagai Ketua Program Studi S2/S3 Statistika ITS yang telah banyak memberikan saran, bimbingan dan masukan agar penulis dapat segera menyelesaikan studi S3.
4. Prof. Drs. Nur Iriawan, M.IKom., Ph.D, sebagai promotor yang selalu saya kagumi, hormati dan dengan penuh keiklasan meluangkan waktu dan tenaga untuk memberikan bimbingan, arahan, saran dan semangat kepada penulis selama menempuh pendidikan Doktor dan terwujudnya buku disertasi ini. Secara tulus penulis sampaikan banyak terimakasih atas bimbingan, petunjuk, arahan, motivasi, ilmu, pelajaran tentang hidup dan doa yang telah diberikan dengan penuh kesabaran selama ini. Semoga Allah SWT selalu memberikan

barokah, nikmat, taufik dan hidayah-Nya kepada bapak sekeluarga. Aamiin 3x YRA.

5. Dr. Irhamah, S.Si, M.Si dan Dr. rer. pol. Heri Kuswanto, S.Si., M.Si, sebagai co-Promotor yang telah memberikan bimbingan, arahan, saran dan semangat kepada penulis selama menempuh pendidikan Doktor dan terwujudnya buku disertasi ini. Semoga Allah SWT selalu memberikan barokah, nikmat, taufik dan hidayah-Nya kepada bapak dan ibu sekeluarga. Aamiin 3x YRA.
6. Dr. Danardono, M.PH, Dr. Brodjol Sutijo Ulama, MS, Dr. Santi Wulan Purnami, S.Si, M.Si, Dr. R. Mohamad Atok, M.Si dan Dr. I Nyoman Latra, MS, sebagai tim penguji eksternal dan internal yang telah banyak memberikan saran dan masukan untuk perbaikan buku disertasi ini.
7. Dr. Dra. Kartika Fithriasari, M.Si dan Dr. Drs. Agus Suharsono, MS, sebagai tim penguji validasi yang telah banyak memberikan saran dan masukan untuk perbaikan buku disertasi ini.
8. Staf pengajar, tata usaha, karyawan dan karyawan di Jurusan Statistika ITS dan UB yang telah banyak memberikan bantuan, dukungan, motivasi dan semangat kepada penulis selama menempuh studi di Program Doktor Statistika ITS ini.
9. Putriku yang cantik dan tersayang Lucha Kamala Putri yang senantiasa selalu mendoakan, mendukung, menyemangati dan mendorong bundanya untuk segera dapat menyelesaikan studi di Program Doktor Statistika ITS dan dukunganmu sungguh sangat berarti bagi bunda.
10. Ibunda terkasih Kuswatun Kasanah dan almarhum Ayahanda terkasih Mochamad Djaswandi, serta saudara-saudaraku tersayang mas Zainal, mbak Tatik, mbak Wiwid dan mbak Yati, yang selalu memberikan dukungan, doa dan kasih sayangnya selama ini.
11. Segenap teman-teman seperjuangan di Program Doktor Jurusan Statistika ITS Surabaya, khususnya kepada Dr. Adji Ahmad Rinaldo Fernandes, S.Si, M.Sc, Achmad Syahrul Choir, S.ST, M.Si dan Dr. Kartika Fitriasari, M.Si yang telah banyak membantu dan selalu menyemangati penulis untuk segera dapat menyelesaikan studi di Program Doktor Statistika ITS ini.
12. Semua pihak yang telah membantu dalam penyelesaian penulisan buku disertasi ini yang tidak dapat disebutkan satu persatu.

Tiada gading yang tak retak, kesempurnaan hanyalah milik Allah SWT semata, penulis sangat menyadari bahwa dalam buku disertasi ini masih diperlukan penyempurnaan dan pengembangan. Oleh karena itu saran dan masukan sangat diharapkan untuk menyempurnakan buku disertasi ini. Penulis sangat berharap buku disertasi ini dapat memberikan manfaat dan memberikan sumbangan dalam menambah wawasan keilmuan bagi semua pihak karena sebaik-baik manusia adalah manusia yang hidupnya selalu bermanfaat bagi orang lain. Aamiin 3x YRA.

Surabaya, 8 Agustus 2017

Penulis



## DAFTAR ISI

	Hal
LEMBAR PENGESAHAN DISERTASI.....	iii
PERNYATAAN.....	v
ABSTRAK.....	vii
ABSTRACT.....	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xv
DAFTAR TABEL.....	xix
DAFTAR GAMBAR.....	xxiii
DAFTAR SYNTAX.....	xxix
DAFTAR LAMPIRAN.....	xxxi
DAFTAR SINGKATAN.....	xxxiii
DAFTAR ALGORITMA.....	xxxv
DAFTAR LAMBANG DAN ARTI.....	xxxvii
<b>BAB I. PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	8
1.3. Tujuan Penelitian .....	8
1.4. Manfaat dan Orisinalitas Penelitian .....	9
1.5. Batasan Penelitian .....	11
1.6. Sistematika Penulisan Disertasi .....	12
<b>BAB II. KAJIAN PUSTAKA DAN DASAR TEORI .....</b>	<b>15</b>
2.1. Data <i>Microarray Chickpea</i> .....	15
2.2. Konsep Pemodelan <i>Ordinary Least Square</i> (OLS) .....	18
2.3. Analisis Bayesian .....	18
2.4. Model <i>Mixture</i> .....	20
2.4.1. Distribusi <i>Mixture</i> Normal .....	21
2.4.2. Model Bayesian <i>Mixture</i> (BMM) .....	22
2.5. Estimasi Parameter .....	24
2.5.1. Fungsi <i>Likelihood</i> .....	24
2.5.2. Distribusi Prior .....	25



2.5.3. Distribusi Bersyarat Penuh .....	27
2.5.4. <i>Directed Acyclic Graph</i> (DAG) Model <i>Mixture</i> .....	32
2.5.5. Metode <i>Markov Chain Monte Carlo</i> (MCMC) .....	33
2.5.6. <i>Metropolis-Hastings</i> .....	37
2.5.7. <i>Reversible Jump</i> .....	38
2.6. Model Bayesian Unimodal Normal (BUM-Normal) .....	41
2.7. Model Bayesian Averaging (BMA) .....	45
2.7.1. Model Bayesian Averaging Normal Unimodal (BMA-Normal Unimodal) .....	49
2.7.2. Seleksi Model dengan Metode Occam's Window .....	50
2.8. Uji Kolmogorov-Smirnov (KS) .....	52
2.9. <i>Highest Probability Distribution</i> (HPD) dengan <i>Credible Interval</i> ( $1 - \alpha$ ) .....	53
BAB III. METODOLOGI PENELITIAN .....	57
3.1. Data Penelitian .....	57
3.2. Program Komputasi dan <i>Software</i> Penelitian .....	59
3.3. Metode Penelitian .....	61
BAB IV. PENGEMBANGAN ALGORITMA RJMCMC DAN SKENARIO PEMBANGKITAN DATA SIMULASI UNTUK DATA <i>CHICKPEA</i> DI INDONESIA .....	73
4.1. Algoritma RJMCMC .....	73
4.2. Algoritma Proses Pendahuluan ( <i>Preliminary Process Algorithm</i> ) .....	75
4.3. Algoritma Proses Pemulusan ( <i>Smoothing Process Algorithm</i> ) .....	76
4.4. Pengembangan Algoritma RJMCMC .....	79
4.5. Skenario Pembangkitan Data Simulasi .....	82
4.6. <i>Goodness of Fit</i> Kolmogorov-Smirnov (KS) .....	83
4.7. Data Simulasi .....	85
4.8. Data <i>Chickpea</i> .....	91

BAB V. BAYESIAN UNIMODAL MODEL DAN BAYESIAN MODEL AVERAGING UNTUK IDENTIFIKASI DATA <i>CHICKPEA</i> .....	107
5.1. Pemodelan dan Algoritma untuk Bayesian Unimodal Model (BUM) Normal .....	107
5.2. Pemodelan dan Algoritma untuk Bayesian Model Averaging (BMA) Normal Unimodal .....	111
5.3. Implementasi Pemodelan untuk Data Simulasi dan Data <i>Chickpea</i> .....	114
5.3.1. Pemodelan BUM-Normal Data Simulasi .....	114
5.3.2. Pemodelan BMA-Normal Unimodal Data Simulasi ...	122
5.3.3. Pemodelan BUM-Normal untuk Data <i>Chickpea</i> .....	125
5.3.4. Pemodelan BMA-Normal Unimodal Data <i>Chickpea</i> ..	130
5.4. Perbandingan Hasil Pemodelan BUM-Normal dan BMA- Normal Unimodal .....	137
BAB VI. IDENTIFIKASI PERBEDAAN EKSPRESI GEN TANAMAN <i>CHICKPEA</i> DI INDONESIA DENGAN PEMODELAN BAYESIAN MIXTURE MODEL .....	139
6.1. Pemodelan dan Algoritma untuk Bayesian <i>Mixture</i> Model (BMM) Normal .....	139
6.2. Implementasi pada Data Simulasi .....	142
6.3. Implementasi pada Data <i>Chickpea</i> .....	161
BAB VII. IDENTIFIKASI PERBEDAAN EKSPRESI GEN TANAMAN <i>CHICKPEA</i> DI INDONESIA DENGAN PEMODELAN BAYESIAN MIXTURE MODEL AVERAGING .....	173
7.1. Pemodelan dan Algoritma untuk Bayesian <i>Mixture</i> Model <i>Averaging</i> (BMMA) Normal .....	173
7.2. Implementasi pada Data Simulasi .....	175
7.3. Implementasi pada Data <i>Chickpea</i> .....	188
7.4. Perbandingan Hasil Pemodelan BMM-Normal dan BMMA-Normal .....	195

BAB VIII. PENUTUP .....	199
8.1. Kesimpulan .....	199
8.2. Saran, Rekomendasi, dan Keberlanjutan Penelitian .....	201
 KARYA-KARYA DI JURNAL DAN <i>PROCEEDING</i> .....	205
DAFTAR PUSTAKA .....	209
LAMPIRAN .....	219
DAFTAR RIWAYAT HIDUP .....	387

## DAFTAR TABEL

Tabel	Hal
Tabel 1.1. Deskripsi Berbagai Penelitian yang telah dilakukan terkait dengan Data Microarray.....	4
Tabel 3.1. Struktur Data Lengkap Data Perbedaan Ekspresi Gen Hasil Penelitian Harijati (2007) .....	60
Tabel 3.2. Kriteria Kelompok ID Gen berdasarkan nilai HPD dalam Pemodelan .....	637
Tabel 4.1. Data Perbedaan Ekspresi Gen pada 15 ID Gen LY-24 dengan Fungsi Ketahanan .....	88
Tabel 4.2. Data Perbedaan Ekspresi Gen pada 11 ID Gen LY-24 dengan Fungsi Energi.....	89
Tabel 4.3. Rata-Rata Data Simulasi (Pembangkitan 1.000 kali) untuk ID Gen LY-24 Fungsi Energi.....	92
Tabel 4.4. Rata-Rata Data Simulasi (Pembangkitan 1.000 kali) untuk ID Gen LY-24 Fungsi Ketahanan .....	923
Tabel 4.5. Statistika Deskriptif Data Simulasi (Pembangkitan 1.000 kali) untuk ID Gen LY-24 Fungsi Ketahanan.....	95
Tabel 4.6. Statistika Deskriptif Data Simulasi (Pembangkitan 1.000 kali) untuk ID Gen LY-24 Fungsi Energi .....	96
Tabel 4.7. Hasil Uji KS Dua Sample Independen antara Data <i>Chickpea</i> dengan Data Simulasi untuk ID Gen Fungsi Ketahanan .....	97
Tabel 4.8. Hasil Uji KS Dua Sample Independen antara Data <i>Chickpea</i> dengan Data Simulasi untuk ID Gen Fungsi Energi .....	97
Tabel 4.9. Statistika Deskriptif Data <i>Chickpea</i> , Perbedaan Ekspresi Gen dari 15 ID Gen LY-24 dengan Fungsi Ketahanan .....	100
Tabel 4.10. Statistika Deskriptif Data <i>Chickpea</i> , Perbedaan Ekspresi Gen dari 11 ID Gen LY-24 dengan Fungsi Energi .....	101
Tabel 4.11. Hasil Uji Jumlah Komponen <i>Mixture</i> dalam Data <i>Chickpea</i> dengan Pengembangan Algoritma RJMCMC.....	104

Tabel 5.1. Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen LS0035 .....	116
Tabel 5.2. <i>Goodness of Fit</i> KS Model BUM-Normal dan Model OLS .....	118
Tabel 5.3. Nilai Batas HPD dengan <i>Credible Interval</i> 95% Model BUM-Normal untuk Data Simulasi .....	119
Tabel 5.4. Hasil Seleksi Model dengan Metode Occam's Window untuk Data Simulasi .....	123
Tabel 5.5. <i>Goodness of Fits</i> KS Model BMA-Normal Unimodal dan Model OLS untuk Data Simulasi .....	125
Tabel 5.6. HPD dengan <i>Credible Interval</i> 95% Model BMA-Normal Unimodal untuk Data Simulasi .....	126
Tabel 5.7. Estimasi Parameter Data <i>Chickpea</i> untuk ID Gen LS0035 .....	126
Tabel 5.8. <i>Goodness of Fits</i> KS Model BUM-Normal dan Model OLSI untuk Data <i>Chickpea</i> .....	129
Tabel 5.9. HPD dengan <i>Credible Interval</i> 95% Model BUM-Normal untuk Data <i>Chickpea</i> .....	130
Tabel 5.10. Hasil Seleksi Model dengan Metode Occam's Window untuk Data <i>Chickpea</i> .....	133
Tabel 5.11. <i>Goodness of Fits</i> KS Model BMA-Normal Unimodal dan Model OLS untuk Data <i>Chickpea</i> .....	133
Tabel 5.12. HPD dengan <i>Credible Interval</i> 95% Model BMA-Normal Unimodal untuk Data <i>Chickpea</i> .....	134
Tabel 5.13. <i>Goodness of Fit</i> KS model BUM-Normal dan BMA-Normal Unimodal untuk Data Simulasi dan Data <i>Chickpea</i> .....	138
Tabel 6.1. Estimasi Parameter Rata-Rata Data Simulasi ID Gen LS0024 dengan Model BMM-Normal Dua Komponen .....	146
Tabel 6.2. Estimasi Parameter Rata-Rata Data Simulasi ID Gen U070 dengan Model BMM-Normal Tiga Komponen .....	149
Tabel 6.3. <i>Goodness of Fit</i> KS Model BMM-Normal Dua dan Tiga Komponen untuk Model OLS Data Simulasi .....	153

Tabel 6.4. HPD dengan <i>Credible Interval</i> 95% untuk Model BMM-Normal dengan Dua dan Tiga Komponen Data Simulasi.....	154
Tabel 6.5. Estimasi Parameter Rata-Rata Data <i>Chickpea</i> untuk ID Gen LS0024 pada Model BMM-Normal Dua Komponen.....	162
Tabel 6.6. Estimasi Parameter Rata-Rata Data <i>Chickpea</i> untuk ID Gen U070 pada Model BMM-Normal Tiga Komponen.....	164
Tabel 6.7. <i>Goodness of Fit</i> KS Model BMM-Normal Dua dan Tiga Komponen untuk Model OLS untuk Data <i>Chickpea</i> .....	170
Tabel 6.8. HPD dengan <i>Credible Interval</i> 95% untuk Model BMM-Normal dengan Dua dan Tiga Komponen Data <i>Chickpea</i> .....	171
Tabel 7.1. Hasil Seleksi Model dengan Metode Occam's Window untuk Data Simulasi.....	178
Tabel 7.2. <i>Goodness of Fit</i> KS Model BMMA-Normal dengan Dua dan Tiga Komponen untuk Model OLS Data Simulasi.....	179
Tabel 7.3. HPD dengan <i>Credible Interval</i> 95% untuk Model BMMA-Normal dengan Dua dan Tiga Komponen Data Simulasi.....	181
Tabel 7.4. Hasil Seleksi Model dengan Metode Occam's Window untuk Data <i>Chickpea</i> .....	192
Tabel 7.5. <i>Goodness of Fit</i> KS Model BMMA-Normal Dua dan Tiga Komponen untuk Model OLS Data <i>Chickpea</i> .....	193
Tabel 7.6. HPD dengan <i>Credible Interval</i> 95% untuk Model BMMA-Normal dengan Dua dan Tiga Komponen Data <i>Chickpea</i> .....	194
Tabel 7.7. <i>Goodness of Fit</i> KS model BMM-Normal dan BMMA-Normal untuk Data Simulasi dan Data <i>Chickpea</i> .....	197
Tabel 7.8. Karya di Jurnal dan <i>Proceeding</i> Internasional.....	205



## DAFTAR GAMBAR

Gambar	Hal
Gambar 1.1. Orisinalitas Penelitian .....	132
Gambar 1.2. Bagan Alir Sistematika Penulisan Disertasi.....	144
Gambar 2.1. Struktur Hirarki dari Prior Model <i>Mixture</i> Normal dengan .....	33
Gambar 2.2. Struktur Hirarki dari Prior Model BUM-Normal dalam DAG .....	46
Gambar 2.3. Ilustrasi Konsep Teori Uji Kolmogorov-Smirnov (KS).....	55
Gambar 2.4. Konsep Densitas Setimbang untuk Pola Data Simetri dan Tidak Simetri, (a) untuk <i>Confidence Interval</i> (b) untuk <i>Credible Interval</i> (Iriawan, 2001) .....	55
Gambar 3.1. Tanaman <i>Chickpea</i> Jenis <i>Lassester</i> pada Umur Muda dan Umur Tua .....	59
Gambar 3.2. Tahapan Identifikasi Ekspresi Gen Jaringan Tanaman <i>Chickpea</i> dengan Teknologi <i>Microarray</i> .....	59
Gambar 3.3. Bagan Alir Metode Penelitian untuk Pembangkitan Data Simulasi .....	69
Gambar 3.4. Bagan Alir Metode Penelitian Secara Umum untuk Pemodelan .....	70
Gambar 3.5. Bagan Alir Metode Penelitian untuk Pemodelan BUM-Normal dan Pemodelan BMA-Normal Unimodal .....	71
Gambar 3.6. Bagan Alir Metode Penelitian untuk Pemodelan BMM-Normal dan Pemodelan BMMA-Normal .....	72
Gambar 4.1. Ilustrasi Konsep Algoritma 4.1 dan Enam Langkah Pertama Algoritma 4.2 untuk Identifikasi Distribusi Data Unimodal.....	78
Gambar 4.2. Plot Vertikal Rata-Rata dari 1.000 Kali Pembangkitan Data Simulasi pada U1, U2 dan U3 untuk setiap ID Gen Fungsi Ketahanan .....	94
Gambar 4.3. Plot Vertikal Rata-Rata dari 1.000 Kali Pembangkitan Data Simulasi pada U1, U2 dan U3 untuk Setiap ID Gen Fungsi Energi.....	94
Gambar 4.4. Plot Vertikal Data <i>Chickpea</i> untuk 15 ID Gen Fungsi Ketahanan .....	102



Gambar 4.5. Plot Vertikal Data <i>Chickpea</i> untuk 15 ID Gen Fungsi Energi .....	102
Gambar 4.6. Dotplot Nilai <i>P-value</i> KS dari 1.000 Kali Pembangkitan Data Simulasi untuk Setiap ID Gen Fungsi Ketahanan .....	103
Gambar 4.7. Dotplot Nilai <i>P-value</i> KS dari 1.000 Kali Pembangkitan Data Simulasi untuk Setiap ID Gen Fungsi Energi .....	103
Gambar 5.1. Ilustrasi Pembentukan HPD dengan <i>Credible Interval</i> 95% .....	109
Gambar 5.2. <i>Graphical Model</i> ( <i>Doodle BUGS</i> ) untuk Model BUM-Normal ....	115
Gambar 5.3. Fungsi Autokorelasi setiap Parameter yang di Estimasi.....	116
Gambar 5.4. Distribusi Kernel setiap Parameter yang di Estimasi .....	116
Gambar 5.5. <i>History</i> Data Bangkitan setiap Parameter yang di Estimasi .....	116
Gambar 5.6. <i>Quantiles</i> setiap Parameter yang di Estimasi untuk ID Gen LS0035.....	117
Gambar 5.7. Bentuk Distribusi BUM-Normal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0035 .....	120
Gambar 5.8. Bentuk Distribusi BUM-Normal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0612 .....	120
Gambar 5.9. Bentuk Distribusi BUM-Normal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0759 .....	120
Gambar 5.10. Bentuk Distribusi BUM-Normal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U278 .....	121
Gambar 5.11. Bentuk Distribusi BUM-Normal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U017 .....	121
Gambar 5.12. Bentuk Distribusi BUM-Normal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0132.....	121
Gambar 5.13. Bentuk Distribusi BUM-Normal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0341 .....	122
Gambar 5.14. Fungsi Autokorelasi setiap Parameter yang di Estimasi.....	126
Gambar 5.15. Distribusi Kernel setiap Parameter yang di Estimasi .....	127

Gambar 5.16. <i>History</i> Data Bangkitan setiap Parameter yang di Estimasi.....	127
Gambar 5.17. Quantiles setiap Parameter yang di Estimasi untuk ID Gen LS0035 .....	128
Gambar 5.18. Bentuk Distribusi BMA-Normal Unimodal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0035 .....	137
Gambar 5.19. Bentuk Distribusi BMA-Normal Unimodal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0612 .....	138
Gambar 5.20. Bentuk Distribusi BMA-Normal Unimodal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0759 .....	138
Gambar 5.21. Bentuk Distribusi BMA-Normal Unimodal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U278 .....	138
Gambar 5.22. Bentuk Distribusi BMA-Normal Unimodal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U017 .....	139
Gambar 5.23. Bentuk Distribusi BMA-Normal Unimodal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0132 .....	139
Gambar 5.24. Bentuk Distribusi BMA-Normal Unimodal: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0341 .....	139
Gambar 6.1. Ilustrasi HPD dengan <i>Credible Interval</i> 95% .....	141
Gambar 6.2. <i>Graphical Model</i> (Doodle BUGS) untuk Model BMM-Normal ...	144
Gambar 6.3. <i>Graphical Model</i> (Doodle BUGS) untuk Model BMM-Normal ...	145
Gambar 6.4. Fungsi Autokorelasi setiap Parameter Model BMM-Normal .....	147
Gambar 6.5. Densitas Kernel setiap Parameter Model BMM-Normal.....	147
Gambar 6.6. <i>History</i> Data Bangkitan setiap Parameter Model BMM-Normal ..	148
Gambar 6.7. <i>Quantiles</i> setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024 .....	149
Gambar 6.8. Fungsi Autokorelasi setiap Parameter Model BMM-Normal .....	149
Gambar 6.9. Densitas Kernel setiap Parameter Model BMM-Normal.....	150
Gambar 6.10. <i>History</i> Data Bangkitan setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070 .....	150

Gambar 6.11. <i>Quantiles</i> setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070 .....	151
Gambar 6.12. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0024.....	154
Gambar 6.13. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0616.....	155
Gambar 6.14. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0159.....	155
Gambar 6.15. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0162.....	155
Gambar 6.16. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0185.....	156
Gambar 6.17. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0081 .....	156
Gambar 6.18. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0752 .....	156
Gambar 6.19. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0688 .....	157
Gambar 6.20. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U273 .....	157
Gambar 6.21. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U070 .....	157
Gambar 6.22. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U325 .....	158
Gambar 6.23. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U385 .....	158
Gambar 6.24. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U447 .....	158
Gambar 6.25. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0426 .....	159
Gambar 6.26. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0534 .....	159

Gambar 6.27. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0702 .....	159
Gambar 6.28. Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA1120 .....	160
Gambar 6.29. Bentuk Distribusi BMM-Normal Tiga Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U001 .....	160
Gambar 6.30. Bentuk Distribusi BMM-Normal Tiga Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0810 .....	160
Gambar 6.31. Fungsi Autokorelasi setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data <i>Chickpea</i> .....	162
Gambar 6.32. Densitas Kernel setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data <i>Chickpea</i> .....	163
Gambar 6.33. <i>History</i> Bangkitan Data Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data <i>Chickpea</i> .....	163
Gambar 6.34. <i>Quantiles</i> setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data <i>Chickpea</i> .....	164
Gambar 6.35. Fungsi Autokorelasi setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070 Data <i>Chickpea</i> .....	165
Gambar 6.36. Densitas Kernel setiap Parameter Model BMM-Normal .....	165
Gambar 6.37. <i>History</i> Bangkitan Data Model BMM-Normal Tiga Komponen untuk ID Gen U070 Data <i>Chickpea</i> .....	166
Gambar 6.38. <i>Quantiles</i> setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070 Data <i>Chickpea</i> .....	167
Gambar 7.1. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0024 .....	182
Gambar 7.2. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0616 .....	182
Gambar 7.3. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0159 .....	182
Gambar 7.4. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0162 .....	183

Gambar 7.5. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0185 .....	183
Gambar 7.6. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0081 .....	183
Gambar 7.7. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0752 .....	184
Gambar 7.8. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen LS0688 .....	184
Gambar 7.9. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U273 .....	184
Gambar 7.10. Bentuk Distribusi BMMA-Normal Tiga Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U070.....	185
Gambar 7.11. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U325.....	185
Gambar 7.12. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U385.....	185
Gambar 7.13. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U447.....	186
Gambar 7.14. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0426.....	186
Gambar 7.15. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0534 .....	186
Gambar 7.16. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0702.....	187
Gambar 7.17. Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA1120.....	187
Gambar 7.18. Bentuk Distribusi BMMA-Normal Tiga Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen U001.....	187
Gambar 7.19. Bentuk Distribusi BMMA-Normal Tiga Komponen: (a) Data <i>Chickpea</i> dan (b) Data Simulasi untuk ID Gen CA0810.....	188

## DAFTAR SYNTAX

<i>Syntax</i>	Hal
<i>Syntax 5.1. BUGS Language</i> untuk Model BUM-Normal.....	115
<i>Syntax 6.1. BUGS Language</i> untuk Model BMM-Normal Dua Komponen .....	145
<i>Syntax 6.2. BUGS Language</i> untuk Model BMM-Normal Tiga Komponen.....	146



## DAFTAR LAMPIRAN

Lampiran		Hal
Lampiran 1.	<i>Code</i> Program R-WinBUGS untuk Membangkitkan Data Simulasi, Membangun Model OLS, BUM-Normal, BMA-Normal, BMM-Normal, dan BMMA-Normal .....	219
Lampiran 2.	Fungsi ID Gen dalam Jaringan Tanaman <i>Chickpea</i> (Harijati, 2007) .....	233
Lampiran 3.	Hasil <i>Running</i> Program BUGS untuk 7 ID Gen pada Model BUM-Normal Data Simulasi .....	235
Lampiran 4.	Hasil <i>Running</i> Program BUGS untuk 7 ID Gen pada Model BUM-Normal Data <i>Chickpea</i> .....	247
Lampiran 5.	Hasil <i>Running</i> Program BUGS untuk 19 ID Gen pada Model BMM-Normal Data Simulasi .....	259
Lampiran 6.	Hasil <i>Running</i> Program BUGS untuk 19 ID Gen pada Model BMM-Normal Data <i>Chickpea</i> .....	313
Lampiran 7.	Code Program R dari Feng (2013) .....	371
Lampiran 8.	Lembar Pernyataan Keaslian Data .....	381
Lampiran 9.	Model Persamaan BMMA dan Persamaan OLS .....	383



## DAFTAR SINGKATAN

BUM	Bayesian Unimodal Model
BMA	Bayesian Model <i>Averaging</i>
BMM	Bayesian <i>Mixture</i> Model
BMMA	Bayesian <i>Mixture</i> Model <i>Averaging</i>
KS	Kolmogorov-Smirnov
MCMC	<i>Markov Chain Monte Carlo</i>
RJMCMC	<i>Reversible Jump Markov Chain Monte Carlo</i>

## DAFTAR ALGORITMA

- Algoritma 2.1. Algoritma MCMC Pendekatan Metode *Gibbs Sampler*
- Algoritma 2.2. Algoritma *Metropolis-Hastings*
- Algoritma 2.3. Algoritma Pembentukan Densitas Setimbang HPD
- Algoritma 4.1. Algoritma RJMCMC
- Algoritma 4.2. Algoritma Proses Pendahuluan (*Preliminary Process Algorithm*)
- Algoritma 4.3. Algoritma Proses Pemulusan (*Smoothing Process Algorithm*)
- Algoritma 4.4. Algoritma Pengembangan Algoritma RJMCMC
- Algoritma 4.5. Algoritma Pembangkitan Data Simulasi
- Algoritma 5.1. Algoritma BUM-Normal
- Algoritma 5.2. Algoritma Seleksi Model Occam's Window
- Algoritma 5.3. Algoritma BMA-Normal Unimodal
- Algoritma 6.1. Algoritma BMM-Normal
- Algoritma 7.1. Algoritma BMM-Normal



## DAFTAR LAMBANG DAN ARTI

$\mathcal{H}'$	Posterior odds suatu model di metode Occam's Window
$\alpha(X,Y)$	Probabilitas penerimaan untuk pergerakan dari $X$ ke $Y$
$\delta_k$	Parameter dari distribusi dirichlet $\mathbf{w}   k$
$B = \sum_{i=1}^n e_i^2$	Jumlah kuadrat <i>error</i>
$D_n$	Nilai selisih terbesar dari $F_n(x_i)$ dengan $\hat{F}(\mathbf{x})$
$\Delta$	Parameter yang akan diprediksi dalam model BMA
$E(\Delta   \mathbf{x})$	Distribusi posterior terboboti dari $\Delta$ di setiap model kombinasi yang mungkin dari data $\mathbf{x}$
$\varepsilon$	Nilai toleransi yang ditetapkan sebesar 0,0999
$f(\mathbf{x})$	Suatu konstanta yang disebut sebagai <i>normalized constant</i> dari data $\mathbf{x}$
$f(\mathbf{x}   \boldsymbol{\theta})$	Fungsi <i>likelihood</i> dari data $\mathbf{x}$ dengan parameter $\boldsymbol{\theta}$ dan $\mathbf{x} = (x_1, x_2, \dots, x_n)$
$f(\mathbf{x}   \boldsymbol{\theta}, \mathbf{w})$	Fungsi probabilitas <i>mixture</i> dari data $\mathbf{x}$ dengan vektor parameter model $\boldsymbol{\theta}$ dan vektor pembobot $\mathbf{w}$
$f(\mathbf{x}   \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$	Fungsi probabilitas <i>mixture</i> normal dari data $\mathbf{x}$ dengan parameter pembobot $\mathbf{w}$ , parameter rata-rata $\boldsymbol{\mu}$ dan parameter simpangan baku $\boldsymbol{\sigma}$
$f(\mathbf{x}   \boldsymbol{\theta}_{z_i})$	Distribusi sub populasi dari data $\mathbf{x}$ jika diberikan nilai $z_i$ dengan parameter $\boldsymbol{\theta}$
$F_n(x_i)$	Fungsi distribusi kumulatif empiris dari data $\mathbf{x}$
$\hat{F}(\mathbf{x})$	Fungsi distribusi kumulatif hipotesis (dugaan) dari data $\mathbf{x}$
$g_j(\mathbf{x}   \theta_j)$	Fungsi probabilitas ke- $j$ dari data $\mathbf{x}$ dengan parameter $\theta_j$ dan $j = 1, 2, \dots, k$
$k$	Jumlah komponen <i>mixture</i>

$\kappa$	Bilangan presisi yang nilainya sebesar $\frac{1}{R^2}$
$\xi$	Nilai <i>midpoint</i> dari $\mu_j$
$l_{mixture}(\boldsymbol{\theta}   \mathbf{x})$	Fungsi <i>likelihood</i> model <i>mixture</i> dari data $\mathbf{x}$ dengan parameter $\boldsymbol{\theta}$
$l_{mixture}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{w}   \mathbf{x})$	Fungsi <i>likelihood</i> model <i>mixture</i> normal dari data $\mathbf{x}$ dengan parameter $\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{w}$
$\boldsymbol{\Lambda}$	Vektor parameter dengan elemen-elemen $\boldsymbol{\Lambda} = (\mathbf{z}, \boldsymbol{\theta}, \mathbf{w}, k)$
$\Lambda_k$	Vektor parameter $\boldsymbol{\Lambda}$ komponen ke- $k$
$\Lambda_{-k}$	Komplemen dari parameter $\Lambda_k$ yaitu vektor parameter $\boldsymbol{\Lambda}$ tanpa memuat parameter dalam komponen ke- $k$
$\text{maks}_l(P(M_l   \mathbf{x}))$	Probabilitas posterior tertinggi pada model $M_l$ dari data $\mathbf{x}$
$\mathbf{M}$	Himpunan model dengan elemen-elemen $\mathbf{M} = \{M_1, M_2, \dots, M_q\}$
$p(\boldsymbol{\theta})$	Distribusi prior bagi parameter $\boldsymbol{\theta}$ dengan elemen-elemen $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$
$p(\boldsymbol{\theta}   \mathbf{x})$	Distribusi prior dari parameter $\boldsymbol{\theta}$ dari data $\mathbf{x}$
$p(k)$	Distribusi prior dari parameter jumlah komponen <i>mixture</i> $k$
$p(k, \mathbf{w}, \mathbf{z}, \boldsymbol{\theta}, \mathbf{x})$	Distribusi posterior gabungan model <i>mixture</i> dari data $\mathbf{x}$
$p(M_k)$	Probabilitas prior jika model $M_k$ benar
$p(\mathbf{x}   \boldsymbol{\theta}, \mathbf{z}, \mathbf{w}, k)$	Fungsi <i>likelihood</i> dari data $\mathbf{x}$ jika diketahui $\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}$ dan $k$ bernilai tertentu
$p(\mathbf{x}   \theta_k, M_k)$	Fungsi <i>likelihood</i> dari data $\mathbf{x}$ pada model $M_k$ dengan parameter $\theta_k$
$p(\mathbf{x}   \boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{z}, \mathbf{w}, k)$	Fungsi <i>likelihood</i> dari data $\mathbf{x}$ jika diketahui $\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{z}, \mathbf{w}$ dan $k$ bernilai tertentu
$p(\mathbf{w}   k)$	Distribusi prior dari parameter pembobot $\mathbf{w}$ dengan syarat $k$ bernilai tertentu

$p(z_i = j)$	Distribusi dari lokasi setiap pengamatan $z_i$ dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, k$
$p(z_i = j   \mathbf{w}, k)$	Distribusi bersyarat penuh dari parameter alokasi $z_i$ bersyarat parameter $\mathbf{w}, k$
$p(\mathbf{z}   \mathbf{w}, k)$	Distribusi prior dari parameter alokasi $\mathbf{z}$ bersyarat $\mathbf{w}$ dan $k$ bernilai tertentu
$p(\mathbf{w}   k)$	Distribusi prior dari parameter pembobot $\mathbf{w}$ bersyarat parameter $k$ bernilai tertentu
$p(\Lambda)$	Distribusi stasioner dari parameter $\Lambda$
$p(\beta   \sigma_j^{-2})$	Distribusi prior dari hiperparameter $\beta$ bersyarat parameter $\sigma_j^{-2}$ bernilai tertentu
$p(\Lambda_k   \Lambda_{-k})$	Distribusi bersyarat penuh dari parameter $\Lambda_k$ bersyarat $\Lambda_{-k}$
$p(\boldsymbol{\mu}   \boldsymbol{\sigma}, \mathbf{z}, \mathbf{w}, k)$	Distribusi prior dari parameter $\boldsymbol{\mu}$ bersyarat $\boldsymbol{\sigma}, \mathbf{z}, \mathbf{w}$ dan $k$ bernilai tertentu
$p(\mu_j   \sigma^2, \mathbf{z}, \mathbf{w}, k)$	Distribusi bersyarat penuh dari parameter lokasi $\mu_j$ bersyarat parameter $\boldsymbol{\sigma}, \mathbf{z}, \mathbf{w}, k$
$p(\boldsymbol{\sigma}   \boldsymbol{\mu}, \mathbf{z}, \mathbf{w}, k)$	Distribusi prior dari parameter $\boldsymbol{\sigma}$ bersyarat $\boldsymbol{\mu}, \mathbf{z}, \mathbf{w}$ dan $k$ bernilai tertentu
$p(\sigma_j^{-2}   \mu_j, \mathbf{z}, \mathbf{w}, k)$	Distribusi bersyarat penuh dari parameter varians $\sigma_j^{-2}$ bersyarat parameter $\mu_j, \mathbf{z}, \mathbf{w}, k$
$p(\boldsymbol{\theta}   \mathbf{z}, \mathbf{w}, k)$	Distribusi prior dari parameter model $\boldsymbol{\theta}$ bersyarat $\mathbf{z}, \mathbf{w}$ dan $k$ bernilai tertentu
$P(M_k   \mathbf{x})$	Distribusi posterior dari model $M_k$ jika diketahui data $\mathbf{x}$
$P(\mathbf{x}   M_k)$	Marginal <i>likelihood</i> dari model $M_k$ untuk data $\mathbf{x}$
$P(X, X) = r(X)$	Probabilitas terjadinya penolakan transisi dari variabel $X$ ke variabel $Y$
$P(X, Y)$	Probabilitas transisi Kernel dari variabel $X$ ke variabel $Y$

$P(\Delta   \mathbf{x})$	Distribusi posterior dari $\Delta$ jika diketahui data $\mathbf{x}$
$p(\theta_k   M_k)$	Probabilitas prior dari $\theta_k$ jika diketahui model $M_k$
$\pi(X)$	Probabilitas transisi Kernel variabel $X$
$\pi(Y)$	Probabilitas transisi Kernel variabel $Y$
$q(X, Y)$	Densitas dari variabel $X$ ke <i>state</i> variabel $Y$ yang akan diajukan untuk mendapatkan rantai Markov berikutnya
$\mathbf{w}$	Vektor parameter pembobot dalam model <i>mixture</i> dengan elemen-elemen $\mathbf{w} = (w_1, w_2, \dots, w_k)$ dengan $\sum_{j=1}^k w_j = 1, j = 1, 2, \dots, k$
$\mathbf{z}$	Parameter alokasi pada model <i>mixture</i>
$\mathcal{G}(M_k)$	Distribusi probabilitas prior dari model $M_k$
$\mathcal{G}(\sigma^2   M_k)$	Distribusi bersyarat dari parameter $\sigma$
$\mathcal{G}(\theta_k   \sigma^2, M_k)$	Distribusi bersyarat dari parameter $\theta$
$q(Y, X)$	Densitas dari variabel $Y$ untuk kembali ke <i>state</i> variabel $X$
$R$	Nilai kisaran ( <i>range</i> ) dari data $\mathbf{x}$
$\Gamma(\alpha, \beta)$	Distribusi gamma dengan parameter $\alpha$ dan $\beta$
$Var(\Delta   \mathbf{x})$	Varians dari $E(\Delta   \mathbf{x})$

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

*Microarray* merupakan suatu teknik analisis untuk memonitor aktifitas ribuan ekspresi gen secara simultan. Dengan kata lain, data ekspresi gen merupakan data yang diperoleh dari percobaan *microarray*. Ekspresi gen merupakan rangkaian proses penerjemahan informasi genetik (dalam bentuk urutan basa pada DNA atau RNA) menjadi protein dan lebih jauh lagi menjadi fenotipe. Informasi yang dibawa bahan genetik tidak bermakna apapun bagi suatu organisme apabila tidak diekspresikan menjadi fenotipe (Baggerly, 2001 dan Knudsen, 2004). *Microarray* adalah salah satu kemajuan terbaru yang digunakan untuk penelitian kanker melalui pendekatan farmakologis untuk mengobati berbagai penyakit. Dalam hal ini, *microarray* membantu dalam menguji tanda-tanda tertentu adanya kanker/tumor pada suatu jaringan tertentu pada manusia, hewan atau tanaman dengan menganalisis ukuran sampel yang telah pernah dicatat sebelumnya ataupun sampel baru (Do dkk., 2004; Sebastiani dkk., 2006; Govindarajan dkk., 2012).

Tantangan utama analisis statistika untuk data hasil percobaan *microarray* adalah ukuran sampel yang tersedia sedikit karena keterbatasan dana, tenaga dan waktu serta keterbatasan ketersediaan ukuran sampel tetapi karakteristik variabel yang terukur sangat besar. Akibatnya, identifikasi distribusi data hasil percobaan *microarray* yang diperoleh akan sangat rumit yang berdistribusi multimodal (Muller dkk., 2002). Hal ini dapat ditunjukkan pada pengamatan perbedaan ekspresi gen dari suatu jaringan tanaman, manusia, ataupun hewan dalam kondisi normal (sehat) ataupun tidak normal (terserang penyakit). Dalam hal ini, ekspresi gen yang terukur sebagai hasil dari percobaan *microarray* berjumlah sangat banyak (ribuan ID gen) walaupun ukuran sampel yang tersedia sangat terbatas. Fenomena ini perlu penanganan khusus terutama untuk kasus perbandingan dua kondisi ekspresi gen yang berbeda karena pada dasarnya kasus perbandingan harus mengikuti sebaran empiris tertentu yang membutuhkan jumlah sampel yang besar (Wit dan McClure, 2004 dan Sebastiani dkk., 2006).



Analisis Bayesian merupakan suatu metode analisis statistika yang tidak mempertimbangkan ukuran sampel dalam data, sehingga analisis ini dapat digunakan untuk data yang memiliki ukuran sampel kecil ataupun besar (Gosh dkk., 2006). Kinerja analisis Bayesian didasarkan langsung pada distribusi dalam data yaitu distribusi peluang posterior yang diperoleh dari perkalian antara fungsi *likelihood* (didasarkan atas data pengamatan yang diperoleh) dan distribusi prior (yang didasarkan atas informasi data masa lalu). Parameter model dalam analisis Bayesian dipandang sebagai variabel random dalam ruang parameter model (Gelman dkk., 1995). Menurut Mengersen (2009), keunggulan analisis Bayesian adalah bahwa metode ini masih tetap dapat memberikan inferensi terhadap suatu parameter yang tidak diketahui berdasarkan distribusi posterior dengan tetap mempertahankan kondisi data apa adanya melalui konsep data *driven* (Iriawan, 2012). Selain itu, analisisnya juga memungkinkan adanya kombinasi yang berbeda-beda dari distribusi prior dan memfasilitasi *updating* iteratif berdasarkan informasi baru sehingga dapat mengatasi ketidakpastian dan kompleksitas model dari data.

Berbagai penelitian yang telah dilakukan sebelumnya, terkait dengan data ekspresi gen sebagai hasil dari percobaan *microarray* adalah penelitian Alon dkk. (1999) yang menggunakan analisis *cluster* untuk mengidentifikasi perbedaan ekspresi gen dan penelitian Harijati (2007) menggunakan Student's *t-test* untuk mengidentifikasi perbedaan ekspresi gen pada jaringan tanaman Kacang Arab atau *Chickpea* (*Cicer Arietinum*). Penelitian Harijati (2007) yang menggunakan data *Chickpea*, diperoleh hasil pengelompokan ID gen berdasarkan perbedaan ekspresi gen yang terjadi, yaitu kelompok ID gen yang masuk dalam *Up-regulated* (ID gen yang tahan dari serangan penyakit), *Regulated* (ID gen yang tidak cukup tahan dari serangan penyakit) dan *Down-regulated* (ID gen yang tidak tahan dari serangan penyakit) dari hasil Student's *t-test* yang diperoleh. Beberapa peneliti lain yang juga menggunakan Student's *t-test* adalah Schena dkk. (1995); Schena dkk. (1996); DeRisi dkk. (1996); Tusher dkk. (2001) dan Lonnstedt dan Speed (2002). Penelitian dari Thomas dkk. (2001), Giles dan Kipling (2003) mengasumsikan bahwa data ekspresi gen mengikuti distribusi Gaussian unimodal. Sementara itu, Baldi dan Long (2001) dan Ibrahim dkk. (2002), keduanya

mengasumsikan bahwa data ekspresi gen mengikuti distribusi log-normal. Sedangkan Chen dkk. (1997) dan Newton dkk. (2001) mengasumsikan bahwa data ekspresi gen mengikuti distribusi gamma. Penelitian Efron dkk. (2001) menggunakan pendekatan analisis Bayes Empiris Nonparametrik untuk data hasil percobaan *microarray*. Penelitian Ibrahim dkk. (2002) juga menggunakan model Bayesian untuk data ekspresi gen hasil percobaan *microarray*. Penelitian Astuti dkk. (2015d; 2016a; 2016b; 2016d; 2016e) telah meneliti tentang data Diabetes Mellitus dengan pendekatan Bayesian dimana distribusi data menggunakan pendekatan normal unimodal. Berdasarkan penelitian-penelitian tersebut dapat ditunjukkan bahwa berbagai pendekatan distribusi unimodal dapat digunakan untuk data ekspresi gen.

Penelitian-penelitian lain yang terkait dengan data *microarray* dengan menggunakan pendekatan model *mixture* adalah Antoniak (1974), Mc. Lachlan dkk. (2002) dan Pan dkk. (2002). Kemudian penelitian Do dkk. (2004) telah mengusulkan *Bayesian Mixture Model* (BMM) untuk mengidentifikasi perbedaan ekspresi gen dengan pendekatan metode peluang Bayesian nonparametrik. Hasil penelitiannya menunjukkan bahwa pendekatan model yang diperoleh sama dengan model pendekatan Bayesian empiris yang diusulkan oleh Efron dkk. (2001). Tetapi metode Bayesian nonparametrik ini memiliki keunggulan dapat mengurangi batasan-batasan penting dari metode Bayes empiris dengan penarikan kesimpulannya tetap mudah dipahami walaupun perhitungannya sedikit rumit dan sulit serta membutuhkan biaya yang lebih dibandingkan dengan metode Bayesian empiris. Dalam penelitian Do dkk. (2004), metode pendugaan parameter tidak dilakukan dengan konsep *averaging* model padahal ada kemungkinan model-model baik yang lain dapat dilibatkan dalam model. Penelitian Astuti dkk. (2016d) telah meneliti tentang pemodelan BMM untuk data kasus Diabetes Mellitus dimana distribusi data menggunakan pendekatan *mixture* normal.

Penelitian Sebastiani dkk. (2006) mengusulkan metode pendekatan *Bayesian Model Averaging* (BMA) untuk menghitung peluang posterior perbedaan ekspresi gen dimana model data ekspresi gen memiliki sebaran prior *hierarchy* yang mengikuti distribusi log-normal dan distribusi gamma. Pendekatan BMA dilakukan karena diduga data ekspresi gen memiliki lebih dari satu

distribusi yang sesuai yaitu distribusi log-normal dan distribusi gamma. BMA melibatkan semua model baik yang mungkin dibentuk dan prediksi model didasarkan atas rata-rata terboboti dari semua model baiknya. Hal ini berarti bahwa estimasi BMA mencakup semua model yang mungkin terbentuk, sehingga diperoleh hasil estimasi yang lebih baik (Madigan dan Raftery, 1994). Astuti dkk. (2015a; 2016a; 2016b; 2016e) telah meneliti tentang pemodelan BMA dan BMMA dengan distribusi data menggunakan pendekatan normal unimodal. Pada penelitian Sebastiani dkk. (2006), distribusi data ekspresi gen tidak diasumsikan berdistribusi *mixture*.

Berdasarkan penelitian-penelitian tentang data *microarray* sebagaimana yang telah diuraikan di atas, maka secara ringkas deskripsi dari masing-masing penelitian tersebut dapat dilihat dalam Tabel 1.1.

**Tabel 1.1.** Deskripsi Berbagai Penelitian yang telah dilakukan terkait dengan Data *Microarray*

No.	Peneliti	Deskripsi	Kelebihan	Kekurangan
1	Alon dkk. (1999)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda menggunakan analisis cluster	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan	Tidak mempertimbangkan adanya indikasi <i>mixture</i> dalam data
2	Harijati (2007)	Identifikasi perbedaan Ekspresi 62 ID gen pada kondisi sehat dan sakit pada jaringan tanaman <i>Chickpea</i> menggunakan Student's <i>t-test</i>	Dapat diketahui ID gen dari jaringan tanaman <i>Chickpea</i> yang termasuk dalam kelompok <i>Up-regulated</i> , <i>Regulated</i> dan <i>Down-regulated</i>	Penggunaan Student's <i>t-test</i> untuk ukuran sampel $n=3$ , tidak mempertimbangkan adanya indikasi <i>mixture</i> dalam data
3.	Schena dkk. (1995); Schena dkk. (1996); DeRisi dkk. (1996); Lonnstedt dan Speed (2002); Tusher dkk. (2003)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda menggunakan Student's <i>t-test</i>	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan	Penggunaan Student's <i>t-test</i> untuk ukuran sampel kecil ( $n=2$ )

**Tabel 1.1.** Lanjutan

No.	Peneliti	Deskripsi	Kelebihan	Kekurangan
4.	Thomas dkk. (2001); Giles dan Kipling (2003)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan asumsi data ekspresi gen memiliki distribusi gaussian	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan asumsi distribusi gaussian	Penggunaan asumsi data ekspresi gen memiliki distribusi tunggal normal
5	Baldi dan Long (2001); Ibrahim dkk. (2002)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan asumsi data ekspresi gen memiliki distribusi log-normal	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan asumsi distribusi log-normal	Penggunaan asumsi data ekspresi gen memiliki distribusi tunggal log-normal
6	Chen dkk. (1997); Newton dkk. (2001)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan asumsi data ekspresi gen memiliki distribusi gamma	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan asumsi distribusi gamma	Penggunaan asumsi data ekspresi gen memiliki distribusi tunggal gamma
7	Efron dkk. (2001)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan <i>empirical Bayes analysis</i>	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan pendekatan Bayes	Penggunaan asumsi <i>free distribution</i> untuk data ekspresi gen
8	Ibrahim, Chen dan Gray (2002)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan Bayesian <i>models</i>	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan pendekatan Bayes	Penggunaan asumsi data ekspresi gen memiliki distribusi tunggal

**Tabel 1.1.** Lanjutan

No.	Peneliti	Deskripsi	Kelebihan	Kekurangan
9	Mc. Lachlan dkk. (2002); Pan dkk. (2002)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan <i>mixture t-distribution model</i> dua komponen	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan pendekatan distribusi <i>mixture</i> .	Penggunaan asumsi distribusi <i>mixture</i> dalam data ekspresi gen dengan metode prediksi posterior bukan BMA
10	Do dkk. (2004)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan Bayesian <i>mixture</i> model	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan mempertimbangkan beberapa distribusi tunggal yang sesuai	Prediksi distribusi posterior bukan dengan metode BMA
11	Sebastiani dkk. (2006)	Identifikasi perbedaan ekspresi gen pada dua kondisi yang berbeda dengan Bayesian model <i>averaging</i>	Dapat diketahui pengelompokan ID gen yang diamati berdasarkan ekspresi gen yang diberikan dengan pendekatan distribusi <i>mixture</i>	Penggunaan asumsi data ekspresi gen bukan distribusi <i>mixture</i>

Berdasarkan dua ide dasar penelitian utama yang digunakan, yaitu penelitian Do dkk. (2004) dan Sebastiani dkk. (2006) maka dalam penelitian ini akan dikaji dan dipelajari serta dikembangkan metode baru sebagai penggabungan dua metode yang telah diusulkan sebelumnya, yaitu Metode Bayesian *Mixture Model Averaging* (BMMA) untuk mengidentifikasi perbedaan ekspresi gen pada data *microarray*. Hasil pengembangan metode baru tersebut diharapkan dapat diketahui metode ini mampu menjawab tantangan analisis statistika untuk data *microarray*. Dalam metode baru ini, data *microarray* diasumsikan mengikuti sebaran *mixture* tertentu dan prediksi parameter model *mixture* digunakan metode BMA, yaitu dengan menetapkan bentuk distribusi peluang posterior *averaging* model *mixture*. Ide penggabungan metode Bayesian *mixture* (BMM) dan Bayesian *averaging* (BMA) ini didasarkan pada tantangan data *Chickpea* kasus di Indonesia, yaitu memiliki indikasi berdistribusi *mixture* dan dapat didekati dengan

berbagai distribusi yang sama-sama baik untuk dimodelkan. Pengembangan metode dengan menggabungkan metode BMM dan BMA ini diharapkan dapat menghasilkan informasi yang lebih tepat dibandingkan dengan metode-metode sebelumnya. Algoritma dan komputasi penggabungan kedua metode Bayesian tersebut akan dibangun melalui Program R dan akan diterapkan pada kasus riil, yaitu kasus data *microarray* Indonesia di bidang Biologi berdasarkan data *Chickpea*, yaitu data ekspresi gen dari jaringan dua jenis tanaman *Chickpea* jenis *Lasseter* (jenis yang rentan terhadap serangan penyakit) dan jenis *FLIP94508C* (jenis yang resisten terhadap serangan penyakit). Jaringan tanaman tersebut diamati pada kondisi sehat dan pada kondisi sakit akibat terserang penyakit *Ascochyta Rabiei*. Umur tanaman *Chickpea* yang diamati adalah pada umur tanaman muda dan tua sedangkan pengamatan untuk waktu serangan penyakit adalah 24 jam (1 hari), 48 jam (2 hari) dan 72 jam (3 hari) setelah serangan penyakit. Pemodelan simulasi dari penggabungan metode ini juga dilakukan untuk mengetahui kehandalan metode yang dikembangkan. Pengembangan metode ini diharapkan dapat memperbaiki kinerja metode-metode sebelumnya yang digunakan, sehingga metode pengembangan ini dapat benar-benar menjawab tantangan analisis statistika untuk data *microarray*. Selain itu, melalui pemodelan dengan metode ini juga dapat diketahui kelompok-kelompok ID gen yang masuk dalam kelompok *Up-regulated*, *Regulated* dan *Down-regulated* yang sangat bermanfaat dalam menemukan tanaman *Chickpea* yang tahan terhadap serangan penyakit.

Tusher dkk. (2001) telah mengembangkan *Software* SAM (*Significance Analysis of Microarrays*) untuk memudahkan pengidentifikasian perbedaan ekspresi gen pada dua kondisi yang berbeda. Mc. Lachlan dkk. (2002) telah mengembangkan *Software* EMMIX-GENE yang dapat digunakan untuk *cluster* ekspresi gen didasarkan pada model *mixture* yaitu *mixture* normal dan *mixture* t. Sebastiani dkk. (2006) menggunakan *software* Badge (*Bayesian Analysis of Differential Gene Expression*), *software* R (version 1,9) yang dihubungkan ke WinBUGS 1,4 melalui *package* R2WinBUGS. Oleh karena *software* R telah banyak digunakan oleh peneliti-peneliti sebelumnya terkait dengan data ekspresi gen, maka pengembangan metode yang akan dilakukan dalam penelitian ini juga

akan diimplementasikan dengan *software* R. Dalam hal ini *software* R bersifat *open source* yang *freeware* serta menyediakan berbagai teknik statistika dan grafik yang luas. Perkembangan *software* R mengalami kemajuan yang pesat dengan fitur-fitur dan kemampuan baru yang muncul secara kontinyu (Maindonald, 2008; Cryer dan Chan, 2008 dan Suhartono, 2009). Beberapa teknik statistika yang disediakan dalam *software* R antara lain analisis Bayesian, pengujian hipotesis statistika, analisis deret waktu, model linear dan nonlinear, analisis lokasi, dan analisis lokasi-waktu.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengembangkan metode analisis bagi data *Chickpea* dengan *Bayesian Mixture Model Averaging* (BMMA) sebagai gabungan metode *Bayesian Mixture* (BMM) dan *Bayesian Averaging* (BMA)?
2. Bagaimana mengembangkan algoritma dan komputasi metode analisis untuk mengolah data *Chickpea* dengan *Bayesian Mixture Model Averaging* (BMMA) melalui Program R-WinBUGS?
3. Bagaimana mengidentifikasi kelompok ID gen data *Chickpea* dalam *Up-regulated*, *Regulated*, dan *Down-regulated* dengan *Bayesian Mixture Model Averaging* (BMMA)?

## 1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan di atas, maka tujuan dari penelitian ini secara umum adalah untuk mengembangkan metode analisis perbedaan ekspresi gen hasil percobaan *microarray* dengan *Bayesian Mixture Model Averaging* (BMMA) sebagai gabungan metode *Bayesian Mixture* (BMM) dan *Bayesian Averaging* (BMA). Metode BMMA ini diharapkan mampu menjawab tantangan analisis data *microarray* dan mampu mengelompokkan ID gen dalam *Up-regulated*, *Regulated*, dan *Down-regulated* yang sangat dibutuhkan informasinya di bidang Biologi dengan tetap mempertahankan karakteristik data aslinya. Algoritma dan komputasi dari metode ini akan dikembangkan dengan

Program R-WinBUGS. Penelitian ini dikembangkan berdasarkan hasil pengembangan dari penelitian Do dkk. (2004) dengan *Bayesian Mixture Model* dan penelitian Sebastiani dkk. (2006) dengan *Bayesian Model Averaging*. Secara lengkap, tujuan dari penelitian ini adalah:

1. Memperoleh metode analisis baru bagi data perbedaan ekspresi gen hasil percobaan *microarray* dengan *Bayesian Mixture Model Averaging* (BMMA).
2. Memperoleh algoritma dan komputasi metode analisis perbedaan ekspresi gen hasil percobaan *microarray* dengan *Bayesian Mixture Model Averaging* (BMMA) melalui Program R-WinBUGS.
3. Memperoleh cara identifikasi kelompok ID gen dalam *Up-regulated*, *Regulated*, dan *Down-regulated* dari data *Chickpea* melalui model *Bayesian Mixture Model Averaging* (BMMA).

#### **1.4. Manfaat dan Orisinalitas Penelitian**

Adapun manfaat penting dari penelitian ini adalah:

##### **1. Pengembangan Teori Statistika**

Memberikan sumbangan pengembangan teori analisis Bayesian untuk data *microarray*, khususnya dalam analisis Bayesian *mixture* model dimana distribusi peluang posterior untuk memprediksi parameter model yang dibangun menggunakan metode Bayesian model *averaging*. Hal ini belum pernah dilakukan oleh peneliti-peneliti sebelumnya sebagaimana yang telah dijabarkan dalam Tabel 1.1 dan Gambar 1.1.

##### **2. Efisiensi Tenaga, Biaya dan Waktu**

Ketersediaan ukuran sampel yang sangat terbatas pada data *microarray* umumnya menimbulkan masalah yang kritis dalam analisis data statistika. Pengembangan metode yang akan diusulkan dapat mengatasi masalah tersebut tanpa harus menambah ukuran sampel dalam data apalagi jika ukuran sampel yang tersedia memang terbatas dimana kesimpulan hasil analisis tetap sah, sehingga hal ini tentu saja memberikan efisiensi tenaga, biaya dan waktu.

##### **3. Ditemukannya tanaman *Chickpea* yang tahan serangan penyakit**



Penerapan metode analisis yang akan dikembangkan pada data kasus riil data *Chickpea* ini mampu memberikan sumbangan penting untuk Balai Penelitian Tanaman (BPT) khususnya tanaman *Chickpea* di Indonesia. Hal ini ditunjukkan bahwa melalui metode analisis ini mampu menemukan kelompok ID gen-ID gen dari tanaman *Chickpea* yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* yang memiliki kemampuan ganas menyerang tanaman *Chickpea*, sehingga produksinya berkurang sampai 50%. Berdasarkan penemuan ini bisa dikembangkan penelitian tentang penyisipan gen yang mampu memperoleh varietas baru dari tanaman *Chickpea* di Indonesia yang tahan terhadap serangan penyakit.

Sebagaimana yang telah disebutkan dalam sub bab sebelumnya bahwa penelitian yang berkaitan dengan data *microarray* telah banyak dilakukan, mulai dari penggunaan *Student t-test* untuk mengidentifikasi perbedaan ekspresi gen yang terindikasi memiliki masalah ukuran sampel yang terbatas dan distribusi data yang kompleks (multimodal), sehingga sangat dimungkinkan hasil analisisnya kurang sesuai dengan harapan. Penggunaan metode Bayesian untuk mengatasi masalah ketersediaan data dan multimodalitas distribusi data juga telah banyak digunakan seperti penelitian dengan Metode *Bayesian Mixture Model* (BMM) untuk data *microarray* oleh Do dkk. (2004) karena terindikasi adanya distribusi *mixture* dalam data *microarray*. Selain itu, secara terpisah juga digunakan metode *Bayesian Model Averaging* (BMA) oleh Sebastiani dkk. (2006) karena dalam data *microarray* dimungkinkan juga beberapa distribusi tunggal yang sesuai, seperti normal, log-normal, gamma, weibull dan lain-lain sehingga metode BMA yang didasarkan atas rata-rata terboboti dari semua kemungkinan distribusi yang sesuai diharapkan lebih cocok untuk memprediksi distribusi peluang posteriornya. Penelitian yang belum dilakukan untuk data *microarray* (Tabel 1.1 dan Gambar 1.1) adalah mengasumsikan data *microarray* memiliki distribusi *mixture* dimana prediksi parameternya menggunakan metode BMA. Hal ini memberikan keuntungan berupa kemampuan menangkap semua kemungkinan distribusi *mixture* yang sesuai untuk data *microarray* dapat berkontribusi dalam prediksi

parameter melalui identifikasi distribusi peluang posterior metode BMA. Dengan menggabungkan dua metode Bayesian ini dihasilkan suatu metode baru yang diberi nama metode *Bayesian Mixture Model Averaging* (BMMA). Metode BMMA ini sangat diharapkan dapat memperoleh hasil kesimpulan yang lebih tepat dibandingkan dengan metode-metode sebelumnya. Metode ini mempertimbangkan semua kemungkinan distribusi *mixture* dalam data dan menggunakan metode pendugaan parameter rata-rata terboboti dari semua kemungkinan distribusi *mixture* yang sesuai dalam data. Oleh karena itu, orisinalitas dalam penelitian ini adalah:

1. Terbentuknya metode analisis baru untuk data *Chickpea* dengan *Bayesian Mixture Model Averaging* (BMMA) yang merupakan gabungan antara metode Bayesian *Mixture* (BMM) dan Bayesian *Averaging* (BMA).
2. Diperolehnya algoritma dan komputasi metode analisis Bayesian *Mixture Model Averaging* (BMMA) dengan Program R-WinBUGS untuk data *Chickpea*.
3. Diperolehnya cara identifikasi kelompok ID gen data *Chickpea* dalam *Up-regulated*, *Regulated*, dan *Down-regulated* melalui model Bayesian *Mixture Model Averaging* (BMMA).

Orisinalitas penelitian dapat dilihat pada sajian Gambar 1.1.

### 1.5. Batasan Penelitian

Untuk mendukung kemudahan dalam pengembangan metode analisis baru di atas, diperlukan beberapa batasan-batasan penelitian sebagai berikut:

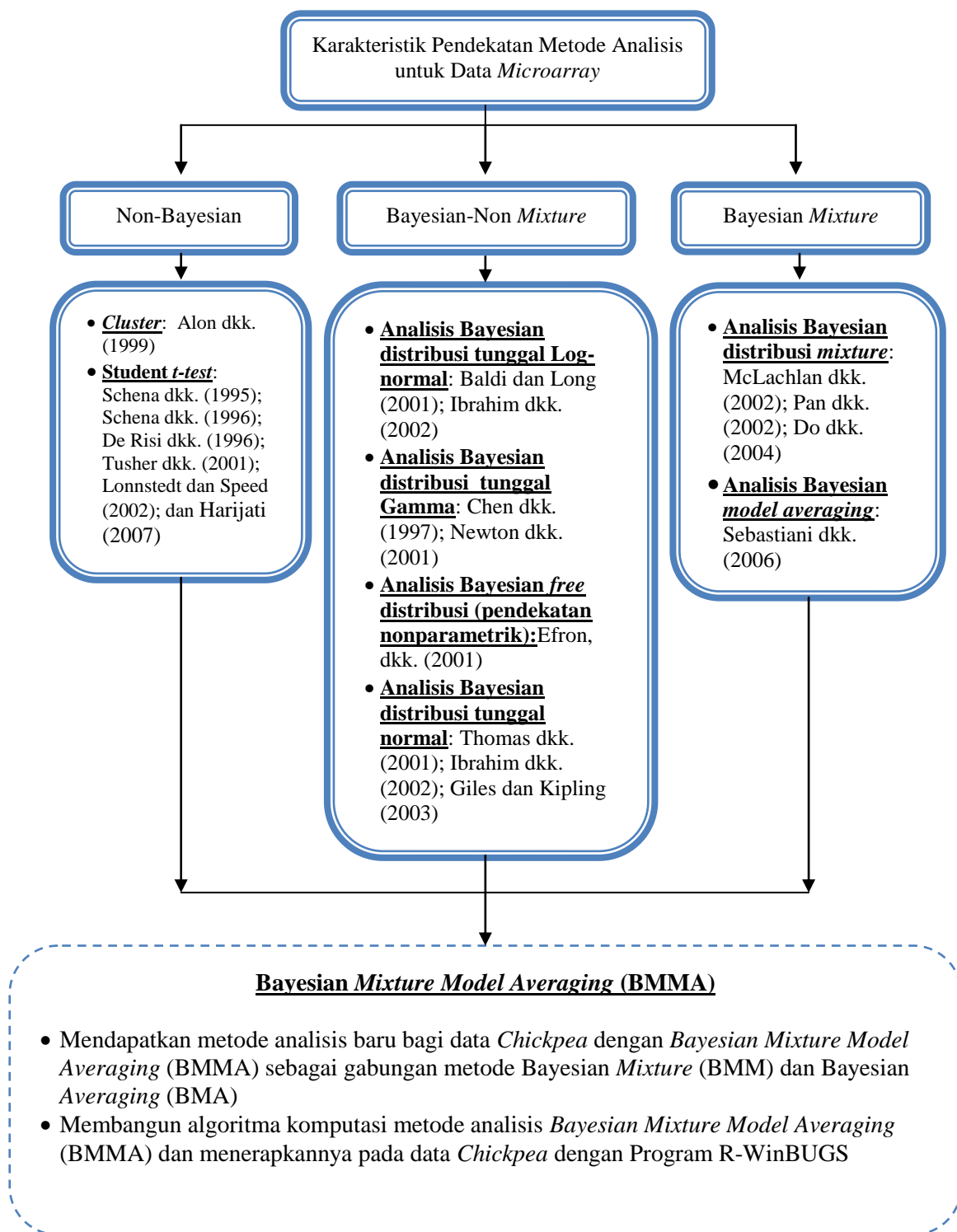
1. Metode seleksi model Occam's Window dari Madigan dan Raftery (1994) digunakan untuk memilih model distribusi *mixture* yang paling sesuai pada pemodelan BMMA. Metode ini dipilih karena banyak digunakan diberbagai penelitian sebelumnya dan menunjukkan hasil identifikasi yang baik.
2. Validasi model BMMA yang terbentuk menggunakan kriteria nilai *goodness of fit* Kolmogorov-Smirnov (KS). Kriteria KS digunakan karena metode ini banyak digunakan pada penelitian-penelitian sebelumnya dan

mudah dalam sisi komputasi khususnya untuk model-model yang kompleks.

3. Metode penentuan jumlah komponen *mixture* yang digunakan adalah *Reversible Jump Markov Chain Monte Carlo* (RJMCMC) karena metode ini sangat fleksibel dalam menentukan jumlah komponen *mixture*.
4. Data kasus riil yang digunakan dalam penelitian ini dikhususkan pada data *microarray* bidang Biologi yaitu data *Chickpea* jenis *Lasseter* muda yang diamati pada dua kondisi sehat dan sakit serta waktu serangan dari penyakit jamur patogen *Ascochyta* Rabiei selama 24 jam (LY-24). ID gen yang diamati perbedaan ekspresinya hanya pada ID gen yang memiliki fungsi ketahanan (15 ID gen) dan fungsi energi (11 ID gen). Hal ini dilakukan karena dua fungsi ID gen tersebut adalah fungsi ID gen yang sangat penting terkait dengan serangan penyakit *Ascochyta* Rabiei dan telah mewakili fenomena yang akan dipelajari dalam penelitian ini.

#### **1.6. Sistematika Penulisan Disertasi**

Tujuan secara umum dari disertasi ini adalah mengembangkan metode Bayesian *Mixture Model Averaging* (BMMA) untuk pemodelan data *Chickpea* di Indonesia. Disertasi ini terdiri dari delapan bab sebagaimana ditunjukkan dalam Gambar 1.2.

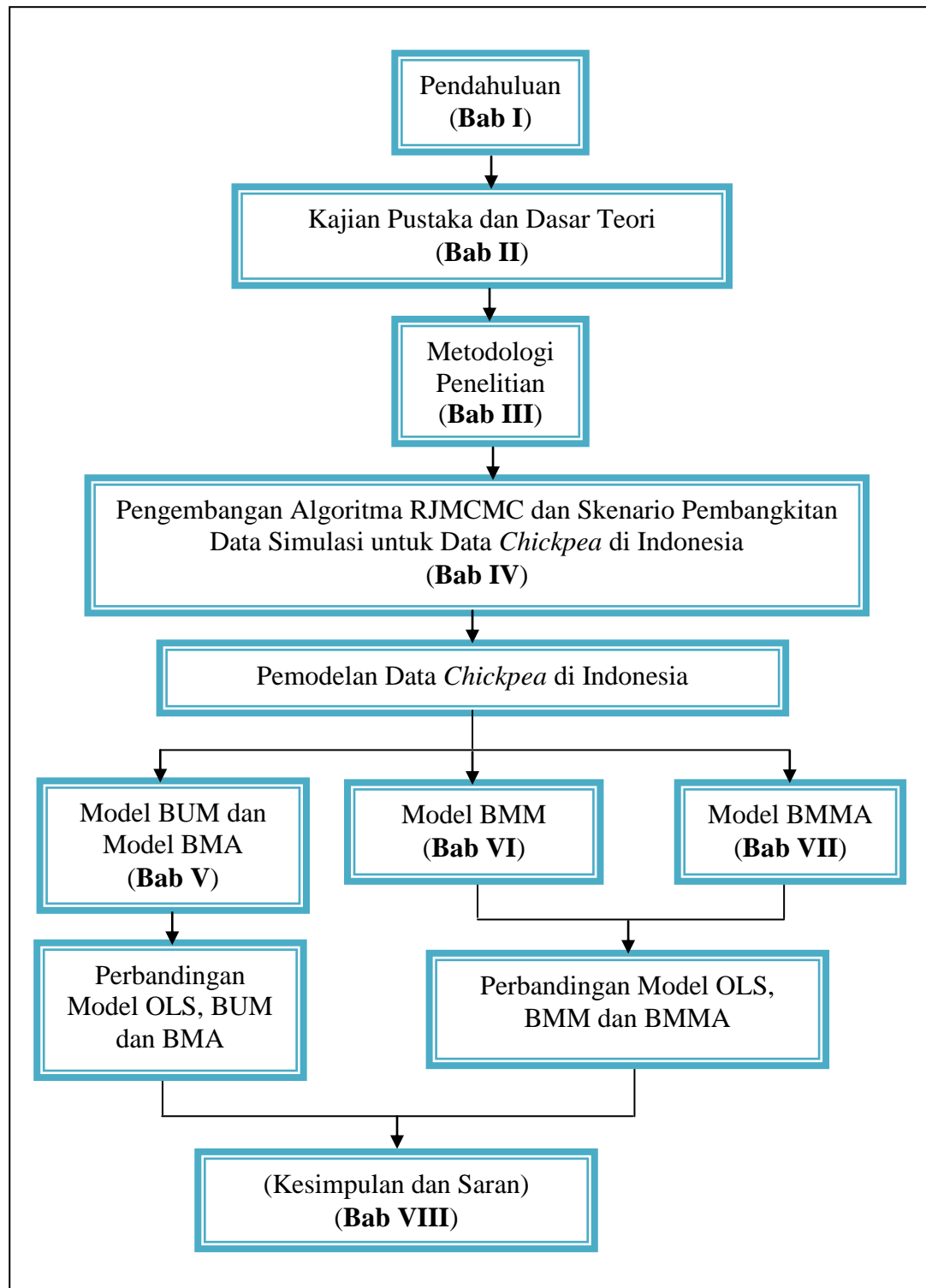


Keterangan:

— = sudah dilakukan

- - - = belum dilakukan (akan diteliti)

**Gambar 1.1.** Orisinalitas Penelitian



Keterangan: BUM : Bayesian Unimodal Model  
 BMA : Bayesian Model *Averaging*  
 BMM : Bayesian *Mixture* Model  
 BMMA : Bayesian *Mixture* Model *Averaging*

**Gambar 1.2.** Sistematika Penulisan Disertasi

## BAB II

### KAJIAN PUSTAKA DAN DASAR TEORI

Dalam bab ini disajikan pembahasan tentang percobaan *microarray* dan tantangan analisis statistika, analisis Bayesian, model *mixture* dan Model *Bayesian Mixture* (BMM), Model Bayesian *Unimodal* (BUM), Model Bayesian *Averaging* (BMA), validasi model dengan *goodness of fit* Kolmogorov-Smirnov (KS), dan *Highest Probability Distribution* (HPD) dengan *Credible Interval*  $(1-\alpha)$ .

#### 2.1. Data *Microarray Chickpea*

Percobaan *microarray* adalah suatu teknik pengumpulan data dengan cara menggunakan *platform* (acuan/pedoman/dasar) yang merupakan hasil duplikasi dari penciri obyek aslinya. Hasil pengukuran dari teknik *microarray* disebut data *microarray* (Knudsen, 2004). Ada beberapa teknologi yang berbeda untuk teknik *microarray*. Sebagai contoh teknik *microarray* yang digunakan dalam penelitian Sebastiani dkk. (2003) yang disebut sebagai teknologi *microarray Synthetic Oligonucleotide*. Teknologi ini melakukan pembagian-pembagian *platform* sebagai cara untuk membagi lokasi-lokasi yang berbeda yang terkait dengan gen yang mengandung banyak duplikasi segmen DNA khusus yang merupakan ciri dari gen tersebut (Duggan dkk., 1999 dan Shoemaker dan Lin, 2005). Segmen DNA khusus tersebut disebut sebagai *Synthetic Oligonucleotide* yang bersifat tetap dan disebut sebagai *probes*. Selanjutnya gen yang diamati dicocokkan dengan *probes* tersebut dan dicatat pasangan *probes*nya dengan gen yang sesuai dan yang tidak sesuai (Nadon dan Shoemaker, 2002; Korenberg, 2007 dan Stafford, 2008).

Data yang diperoleh dari percobaan *microarray* memiliki ciri-ciri sebagai berikut (Muller dkk., 2002):

1. Distribusi dasar dari data yang dihasilkan dengan teknik *microarray* memiliki sifat sangat kompleks (multimodal).
2. Ukuran sampel yang dapat diamati sangat terbatas (sedikit), karena keterbatasan dana, tenaga dan waktu serta ketersediaan ukuran sampel.

3. Karakteristik variabel yang dapat diamati sangat besar, yaitu bisa mencapai puluhan ribu karakteristik dalam setiap percobaan.

Berdasarkan ciri-ciri yang dimiliki oleh data *microarray*, maka dalam penanganannya memerlukan tindakan khusus dan bersifat kompleks karena pada dasarnya metode statistika parametrik memerlukan ukuran sampel yang cukup besar untuk dapat memenuhi derajat bebas pemodelannya. Apabila pemenuhan ukuran sampel ini terlanggar maka kesimpulan dari hasil analisisnya tidak dapat dipertanggungjawabkan (Sebastiani dkk., 2006).

Data ekspresi gen merupakan data hasil percobaan *microarray*, sehingga data ekspresi gen memiliki ciri data *microarray*, yaitu diukur dari sejumlah sampel jaringan yang terbatas (sedikit) tetapi ekspresi gen yang dapat diamati pada setiap percobaan bisa mencapai puluhan ribu karakteristik. Sebagaimana data ekspresi gen yang digunakan dalam penelitian Do dkk. (2004) dengan ukuran sampelnya hanya sebanyak 44 pasien namun data ekspresi gen yang teramati sebanyak 6.500 dari jaringan usus besar manusia yang terkena tumor dan yang tidak terkena tumor. Sedangkan data penelitian dari Sebastiani dkk. (2006), data ekspresi gen yang teramati sebanyak 14.000 dari jaringan sel yang sehat dan yang terkena tumor dengan jumlah sampel hanya sebanyak lima pasien. Berdasarkan pengalaman-pengalaman para peneliti yang menggunakan data ekspresi gen dapat diketahui bahwa data ekspresi gen memiliki ciri-ciri percobaan *microarray*. Oleh sebab itu data ekspresi gen membutuhkan penanganan khusus, baik data ekspresi gen yang diamati pada jaringan manusia, hewan dan tanaman. Penciptaan metode khusus ini dapat mengakomodasi kondisi data yang bersifat sangat kompleks yang pada akhirnya dapat memberikan kesimpulan hasil analisis yang tepat dan valid (Antoniak, 1974 dan Dudoit dkk., 2001).

Data ekspresi gen sebagai hasil dari percobaan *microarray* dari penelitian Harijati (2007) merupakan perbedaan ekspresi gen dari jaringan tanaman *Chickpea* jenis *Lasseter* dan *FLIP94508C*. Jenis *Lasseter* dipilih karena jenis ini tidak tahan (*susceptible*) terhadap serangan penyakit. Sedangkan jenis *FLIP94508C* merupakan jenis yang tahan (*resistant*) terhadap serangan penyakit. Kedua jenis yang saling bertolak belakang tersebut dipilih agar dapat diketahui efektifitas serangan penyakit yang diberikan. Selain itu, juga ingin diketahui

kedua jenis tanaman *Chickpea* ini dalam menghadapi serangan penyakit berdasarkan perubahan ekspresi gen pada jaringannya. Tanaman *Chickpea* dipilih dalam penelitian Harijati (2007) karena tanaman ini merupakan *legume* ketiga yang banyak dibutuhkan di dunia, yaitu tanaman yang merupakan anggota dari famili *leguminosae (fabaceae)* yang buahnya berbentuk polong, bijinya kaya protein, akarnya mengandung bakteri simbiosis pengikat nitrogen dari udara.

*Treatment* terhadap tanaman *Chickpea* yang dilakukan ada dua, yaitu *treatment* pertama, tanaman dibiarkan tumbuh dengan kondisi sehat (*uninoculated*) dan *treatment* kedua, tanaman yang diinokulasi penyakit *Ascochyta Rabiei* sehingga kondisinya sakit (*inoculated*). Kondisi tanaman sakit dibuat dengan cara menyemprotkan air suling yang mengandung suspensi *codinia* ( $1 \times 10^5 \text{ conidia.ml}^{-1}$ ) dan disimpan dalam ruang lembab selama 24 jam. Untuk kondisi tanaman sehat dibuat dengan cara menyemprotkan air suling saja. Serangan terhadap tanaman dipilih jamur patogen *Ascochyta Rabiei*, karena jamur patogen *Ascochyta Rabiei* merupakan penyakit ganas pada tanaman yang dapat menurunkan produksi tanaman hingga 50%.

Pengamatan ekspresi gen terhadap jaringan tanaman *Chickpea* pada masing-masing *treatment* dan jenis tanaman dilakukan pengulangan sebanyak tiga kali dimana antar pengulangan dilakukan secara independen. Artinya antar pengulangan menggunakan tanaman yang berbeda. Selain itu, pengamatan dilakukan pada tiga kelompok waktu yang berbeda, yaitu 24 jam setelah inokulasi atau 24 hpi (*hour post inoculation*), 48 hpi dan 72 hpi. Efek serangan penyakit pada penelitian ini mengakibatkan kondisi tanaman menjadi sakit. Hal ini dilakukan dengan tujuan agar dapat diketahui jenis ID gen tanaman *Chickpea* yang tahan maupun yang tidak tahan terhadap serangan jamur patogen *Ascochyta Rabiei*. Informasi ini sangat dibutuhkan agar dapat dilakukan penyisipan gen yang tahan terhadap serangan jamur patogen *Ascochyta Rabiei* pada tanaman *Chickpea*. Keberhasilan penyisipan gen ini tentu sangat bermanfaat dalam meningkatkan produksi tanaman *Chickpea* di dunia. Hasil pengamatan ekspresi gen jaringan tanaman *Chickpea* ditunjukkan adanya ribuan ID gen yang teramati dalam data *Chickpea* pada penelitian Harijati (2007).



## 2.2. Konsep Pemodelan *Ordinary Least Square* (OLS)

Konsep pemodelan OLS adalah membentuk model dengan prinsip pendekatan OLS, yaitu mengestimasi parameter dari model yang dispesifikasikan dengan meminimumkan jumlah kuadrat *error*,  $\sum_{i=1}^n e_i^2$ . *Error* yang disimbolkan dengan  $e$  didefinisikan sebagai perbedaan antara nilai sebenarnya dan nilai yang diestimasi dari model. Setiap titik data memiliki satu *error* dan jika *error* ini dijumlahkan ataupun di rata-ratakan maka akan bernilai nol (Walpole, 1993; Darlington dan Hayes, 2017).

Konsep OLS dapat dijabarkan sebagai berikut: apabila ada data  $\mathbf{X}$  yang dimodelkan mengikuti distribusi normal dan diketahui bahwa  $B = \sum_{i=1}^n e_i^2$  serta  $\theta$  adalah parameter dari model normal, maka untuk mengestimasi parameter  $\theta$  dengan metode OLS perlu dibentuk persamaan-persamaan normal. Persamaan-persamaan normal itu dapat diperoleh melalui operasi matematika diferensiasi pertama dari persamaan  $B = \sum_{i=1}^n e_i^2$  terhadap masing-masing parameter  $\theta$  yang akan diestimasi dan hasilnya disamakan dengan nol, yaitu  $\frac{\partial B}{\partial \theta} = 0$  (Walpole, 1993; Darlington dan Hayes, 2017).

Jika konsep OLS diimplementasikan pada model normal, maka ada dua parameter dari distribusi normal yang akan diestimasi, yaitu  $\mu$  dan  $\sigma$ , sehingga berdasarkan konsep OLS tersebut akan terbentuk dua persamaan normal dari kedua parameter tersebut, yaitu:

$$(1) \frac{\partial B}{\partial \mu} = 0, \text{ dan } (2) \frac{\partial B}{\partial \sigma} = 0.$$

Berdasarkan kedua persamaan normal tersebut akan diperoleh nilai estimasi untuk parameter  $\mu$  dan  $\sigma$ , yaitu  $\hat{\mu}$  dan  $\hat{\sigma}$ .

## 2.3. Analisis Bayesian

Konsep analisis Bayesian didasarkan dari konsep pemikiran Teorema Bayes yang ditemukan oleh Thomas Bayes pada tahun 1702-1761. Dalam analisis

ini parameter model  $\theta$  dipandang sebagai variabel random dalam ruang parameter  $\Omega$ . Dalam konsep tersebut dijelaskan bahwa analisis Bayesian merupakan suatu metode analisis statistika yang didasarkan pada model distribusi probabilitas posterior dengan struktur sebagai perpaduan dari dua informasi, yaitu informasi data masa lalu dan data pengamatan saat penelitian (sampel). Data masa lalu sebagai informasi prior, sedangkan data pengamatan sebagai penyusun likelihood yang sekaligus sebagai bahan untuk meng-*update* informasi prior (Box dan Tiao, 1973; Zellner, 1971; Gelman dkk., 1995; Iriawan, 2003a dan Congdon, 2006).

Konsep analisis Bayesian dapat dijelaskan sebagai berikut: dimisalkan jika ada data pengamatan  $\mathbf{x}$  yang memiliki fungsi *likelihood*  $f(\mathbf{x}|\theta)$ , maka informasi yang diketahui tentang parameter  $\theta$  sebelum pengamatan itu dilakukan disebut sebagai prior  $\theta$ , yaitu  $p(\theta)$ . Selanjutnya, untuk menentukan distribusi probabilitas posterior dari  $\theta$ , yaitu  $p(\theta|\mathbf{x})$  didasarkan pada aturan probabilitas dalam Teorema Bayes (Ghosh dkk., 2006 dan Dinar, 2009) sebagai berikut:

$$p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)p(\theta)}{f(\mathbf{x})} \quad (2.1)$$

dimana

$$f(\mathbf{x}) = \begin{cases} \int_{\theta \in R} f(\mathbf{x}|\theta)p(\theta)d\theta, & \text{jika } \theta \text{ kontinyu,} \\ \sum_{\theta \in B} p(\mathbf{x}|\theta)p(\theta), & \text{jika } \theta \text{ diskrit,} \end{cases}$$

dengan  $f(\mathbf{x})$  adalah suatu konstanta yang disebut sebagai *normalized constant* (Gelman dkk., 1995), sehingga persamaan (2.1) dapat ditulis dalam bentuk proporsional sebagai:

$$p(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)p(\theta). \quad (2.2)$$

$$\text{posterior} \propto (\text{fungsi } \textit{likelihood}) \times (\text{prior})$$

Berdasarkan persamaan (2.2) diperoleh informasi bahwa probabilitas posterior proporsional terhadap perkalian antara fungsi *likelihood* dan probabilitas prior dari parameter model. Hal ini berarti bahwa informasi prior akan di *update* menggunakan informasi sampel dalam *likelihood* data untuk memperoleh

informasi posterior yang akan digunakan dalam pengambilan keputusan (Iriawan, 2003a).

## 2.4. Model *Mixture*

Model *mixture* merupakan suatu model khusus untuk data yang memiliki sifat multimodal, yaitu data yang di dalamnya memiliki susunan sub-sub populasi atau group dimana setiap sub populasi merupakan komponen penyusun dari model *mixture* yang memiliki proporsi yang berbeda-beda. Model *mixture* disebut model khusus karena model ini mampu menggabungkan data dengan tetap mempertahankan sifat-sifat/karakteristik data aslinya (Mc. Lachlan dan Basford, 1988; Escobar dan West, 1995; Gelman dkk., 1995 dan Astuti, 2006). Selain itu, model *mixture* mampu menggabungkan sejumlah komponen yang berasal dari distribusi yang berbeda-beda sehingga dapat memberikan gambaran (deskripsi) yang lebih realistis dari karakteristik data aslinya. Distribusi *mixture* menyediakan kerangka parametrik yang fleksibel dalam pemodelan dan analisis statistika (Robert, 1996 dan Marin dkk., 2005).

Menurut Mc. Lachlan dan Basford (1988), Gelman dkk. (1995), Iriawan (2001a), dan Prihartanti (2002), fungsi probabilitas *mixture* yang merupakan sebuah model probabilistik dari suatu pengamatan  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  yang diambil dari sejumlah  $k$  sub populasi adalah sebagai berikut:

$$f(\mathbf{x} | \boldsymbol{\theta}, \mathbf{w}) = w_1 g_1(\mathbf{x} | \theta_1) + \dots + w_k g_k(\mathbf{x} | \theta_k) = \sum_{j=1}^k w_j g_j(\mathbf{x} | \theta_j), \quad (2.3)$$

dimana:

$f(\mathbf{x} | \boldsymbol{\theta}, \mathbf{w})$  = fungsi probabilitas *mixture* dari data  $\mathbf{x}$  dengan vektor parameter model  $\boldsymbol{\theta}$  dan vektor pembobot  $\mathbf{w}$ ,

$g_j(\mathbf{x} | \theta_j)$  = fungsi probabilitas ke-  $j$ ,  $j = 1, 2, \dots, k$  dengan parameter  $\theta_j$  yang merupakan vektor parameter yang karakteristiknya tergantung dari bentuk distribusi  $g_j$  masing-masing komponen dalam model *mixture*, dan

$\mathbf{w}$  = vektor parameter proporsi (pembobot) dari model *mixture* dengan elemen-elemen  $w_1, w_2, \dots, w_j$  dimana  $0 < w_j < 1, \forall j$  dengan

$$j = 1, 2, \dots, k \text{ dan } \sum_{j=1}^k w_j = 1.$$

Model *mixture* yang dijabarkan dalam persamaan (2.3) berlaku untuk pemodelan *mixture* dengan banyaknya komponen  $k$  tertentu dimana hal ini disebut sebagai *finite mixture model* (Astuti, 2006).

#### 2.4.1. Distribusi *Mixture Normal*

Apabila diketahui sejumlah  $k$  sub populasi dengan masing-masing sub populasi memiliki distribusi normal, maka persamaan (2.3) dapat dijabarkan menjadi distribusi *mixture* normal sebagaimana disajikan dalam persamaan (2.4) (Mc. Lachlan dan Basford, 1988 dan Gelman dkk., 1995). Parameter  $\theta_j = (\mu_j, \sigma_j^2)$ ,  $j = 1, 2, \dots, k$  dan  $g_j(\mathbf{x} | \theta_j)$  adalah  $N(\mu_j, \sigma_j^2)$ .

$$f(\mathbf{x} | \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = w_1 g_1(\mathbf{x} | \mu_1, \sigma_1^2) + \dots + w_k g_k(\mathbf{x} | \mu_k, \sigma_k^2) = \sum_{j=1}^k w_j g_j(\mathbf{x} | \mu_j, \sigma_j^2), \quad (2.4)$$

dimana:  $-\infty < \mu_j < \infty$ ,  $\sigma_j^2 > 0$ , dan

$f(\mathbf{x} | \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$  = fungsi probabilitas *mixture* normal dari data  $\mathbf{x}$  dengan parameter pembobot  $\mathbf{w}$ , parameter rata-rata  $\boldsymbol{\mu}$  dan parameter simpangan baku  $\boldsymbol{\sigma}$ . Parameter  $\mathbf{w}$  sebagai pembobot memiliki elemen-elemen  $w_1, w_2, w_3$  serta

$g_j(\mathbf{x} | \mu_j, \sigma_j^2)$  = fungsi probabilitas normal ke- $j$  dari sebanyak  $k$  komponen penyusun model *mixture* dengan parameter  $(\mu_j, \sigma_j^2)$ .

Distribusi di masing-masing sub populasi pada persamaan (2.4) dapat dijabarkan dalam persamaan (2.5).

$$g_j(\mathbf{x} | \mu_j, \sigma_j^2) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp \left\{ -\frac{(\mathbf{x} - \mu_j)^2}{2\sigma_j^2} \right\}, \quad j = 1, 2, \dots, k. \quad (2.5)$$

Jika jumlah komponen *mixture* sebanyak dua atau tiga, maka distribusi *mixture* normal dalam persamaan (2.4) dapat dijabarkan dalam persamaan (2.6) dan persamaan (2.7).

$$f(\mathbf{x} | \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = w_1 \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu_1)^2}{2\sigma_1^2} \right\} + w_2 \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu_2)^2}{2\sigma_2^2} \right\}, \quad (2.6)$$

dengan  $-\infty < \mu_1 < \infty$ ,  $-\infty < \mu_2 < \infty$ ,  $\sigma_1^2 > 0$ ,  $\sigma_2^2 > 0$ ,  $0 \leq w_j \leq 1$ ,  $\sum_{j=1}^2 w_j = 1$

$$f(\mathbf{x} | \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = w_1 \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu_1)^2}{2\sigma_1^2} \right\} + w_2 \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu_2)^2}{2\sigma_2^2} \right\} + w_3 \frac{1}{\sigma_3 \sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu_3)^2}{2\sigma_3^2} \right\}. \quad (2.7)$$

dengan  $-\infty < \mu_1 < \infty$ ,  $-\infty < \mu_2 < \infty$ ,  $-\infty < \mu_3 < \infty$ ,  $\sigma_1^2 > 0$ ,  $\sigma_2^2 > 0$ ,  $\sigma_3^2 > 0$ ,  $0 \leq w_j \leq 1$ ,  $\sum_{j=1}^3 w_j = 1$

#### 2.4.2. Model Bayesian Mixture (BMM)

Analisis Bayesian pada model *mixture* memandang bahwa semua parameter dalam model sebagai variabel random yang memiliki distribusi prior tertentu, sehingga analisis ini membutuhkan spesifikasi distribusi prior untuk setiap parameter dalam model (Aitkin, 2001). Hal ini membentuk susunan hirarki pembentuk model sebagai berikut:

- Banyaknya komponen  $k$  dalam model,  $k > 1$
- Proporsi pembobot  $\mathbf{w} = (w_1, w_2, \dots, w_k)$  dimana  $0 \leq w_j \leq 1$ ,  $\forall j$  dan

$$\sum_{j=1}^k w_j = 1, j = 1, 2, \dots, k$$

- Parameter-parameter spesifik dari komponen-komponen  $\theta_j$

Menurut Richardson dan Green (1997), dalam menyusun model *mixture* memandang bahwa setiap pengamatan  $x_i$  akan menjadi anggota dari salah satu sub populasi yang belum diketahui. Jika alokasi setiap pengamatan di setiap sub populasi dari model *mixture* pada persamaan (2.3) dinotasikan dengan  $\mathbf{z}$ , maka proporsi alokasi setiap pengamatan  $z_i$  ditentukan dari distribusi berikut:

$$p(z_i = j) = w_j, i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, k. \quad (2.8)$$

Berdasarkan persamaan (2.8), jika diberikan nilai  $z_i$ , maka sesuai dengan persamaan (2.3) data pengamatan  $x_i$  berasal dari sub populasi yang berdistribusi berikut:

$$x_i | z_i \sim g_j(\mathbf{x} | \boldsymbol{\theta}_{z_i}), i = 1, 2, \dots, n. \quad (2.9)$$

Dengan demikian, distribusi posterior gabungan yang dihasilkan oleh model *mixture* akan berbentuk distribusi gabungan dari semua variabel pada model *mixture* sebagaimana yang dijabarkan di persamaan (2.10).

$$p(k, \mathbf{w}, \mathbf{z}, \boldsymbol{\theta}, \mathbf{x}) = p(k)p(\mathbf{w} | k)p(\mathbf{z} | \mathbf{w}, k)p(\boldsymbol{\theta} | \mathbf{z}, \mathbf{w}, k)p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}, \mathbf{w}, k), \quad (2.10)$$

dimana:

- $p(k, \mathbf{w}, \mathbf{z}, \boldsymbol{\theta}, \mathbf{x})$  = distribusi posterior gabungan model *mixture* dengan banyaknya komponen *mixture*  $k$  dari data  $\mathbf{x}$  dengan parameter model  $\boldsymbol{\theta}$ , pembobot  $\mathbf{w}$  dan parameter alokasi  $\mathbf{z}$ ,
- $p(k)$  = distribusi prior dari parameter jumlah komponen *mixture* sebanyak  $k$ ,
- $p(\mathbf{w} | k)$  = distribusi prior dari parameter pembobot  $\mathbf{w}$  dengan syarat  $k$  bernilai tertentu,
- $p(\mathbf{z} | \mathbf{w}, k)$  = distribusi prior dari parameter alokasi  $\mathbf{z}$  bersyarat  $\mathbf{w}$  dan  $k$  bernilai tertentu,
- $p(\boldsymbol{\theta} | \mathbf{z}, \mathbf{w}, k)$  = distribusi prior dari parameter model  $\boldsymbol{\theta}$  bersyarat  $\mathbf{z}$ ,  $\mathbf{w}$  dan  $k$  bernilai tertentu, dan
- $p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{z}, \mathbf{w}, k)$  = fungsi *likelihood* dari data  $\mathbf{x}$  diketahui  $\boldsymbol{\theta}$ ,  $\mathbf{z}$ ,  $\mathbf{w}$  dan  $k$  bernilai tertentu.

Apabila persamaan (2.10) diaplikasikan pada distribusi *mixture* normal dengan parameter model  $\boldsymbol{\mu}$  dan  $\boldsymbol{\sigma}$ , maka distribusi posterior gabungan untuk semua parameter dalam model *mixture* normal dapat dijabarkan dalam persamaan (2.11).

$$p(k, \mathbf{w}, \mathbf{z}, \boldsymbol{\theta}, \mathbf{x}) = p(k)p(\mathbf{w} | k)p(\mathbf{z} | \mathbf{w}, k)p(\boldsymbol{\mu} | \mathbf{z}, \mathbf{w}, k)p(\boldsymbol{\sigma} | \boldsymbol{\mu}, \mathbf{z}, \mathbf{w}, k)p(\mathbf{x} | \boldsymbol{\sigma}, \boldsymbol{\mu}, \mathbf{z}, \mathbf{w}, k), \quad (2.11)$$

dimana:

- $p(k, \mathbf{w}, \mathbf{z}, \boldsymbol{\theta}, \mathbf{x})$  = distribusi posterior gabungan model *mixture* dengan banyaknya komponen *mixture*  $k$  dari data  $\mathbf{x}$  dengan parameter model  $\boldsymbol{\theta}$  yang berupa  $\boldsymbol{\mu}$  dan  $\boldsymbol{\sigma}$ , pembobot  $\mathbf{w}$  dan parameter alokasi  $\mathbf{z}$ ,
- $p(k)$  = distribusi prior dari parameter banyaknya komponen *mixture*  $k$ ,
- $p(\mathbf{w} | k)$  = distribusi prior dari parameter pembobot  $\mathbf{w}$  dengan syarat  $k$  bernilai tertentu,
- $p(\mathbf{z} | \mathbf{w}, k)$  = distribusi prior dari parameter  $\mathbf{z}$  bersyarat  $\mathbf{w}$  dan  $k$  bernilai tertentu,
- $p(\boldsymbol{\mu} | \boldsymbol{\sigma}, \mathbf{z}, \mathbf{w}, k)$  = distribusi prior dari parameter  $\boldsymbol{\mu}$  bersyarat  $\boldsymbol{\sigma}$ ,  $\mathbf{z}$ ,  $\mathbf{w}$  dan  $k$  bernilai tertentu,

$p(\boldsymbol{\sigma}|\boldsymbol{\mu},\mathbf{z},\mathbf{w},k)$  = distribusi prior dari parameter  $\boldsymbol{\sigma}$  bersyarat  $\boldsymbol{\mu}$ ,  $\mathbf{z}$ ,  $\mathbf{w}$  dan  $k$  bernilai tertentu, dan  
 $p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma},\mathbf{z},\mathbf{w},k)$  = fungsi *likelihood* dari data  $\mathbf{x}$  bersyarat  $\boldsymbol{\mu}$ ,  $\boldsymbol{\sigma}$ ,  $\mathbf{z}$ ,  $\mathbf{w}$  dan  $k$  bernilai tertentu.

Proses selanjutnya adalah melakukan estimasi setiap parameter di persamaan (2.11) melalui pembentukan distribusi bersyarat penuh dari setiap parameter model. Pembentukan distribusi bersyarat penuh dari setiap parameter model dilakukan untuk memudahkan estimasi setiap parameter model *mixture* yang seringkali cukup kompleks dan rumit sehingga tidak mudah untuk diestimasi (Richardson dan Green, 1997 dan Dinar, 2009).

## 2.5. Estimasi Parameter

Estimasi parameter untuk persamaan (2.11) dilakukan dengan membentuk distribusi bersyarat penuh untuk masing-masing parameter dalam model. Untuk membentuk distribusi bersyarat penuh harus diketahui fungsi *likelihood* gabungan dan distribusi prior dari masing-masing parameter model yang akan diestimasi.

### 2.5.1. Fungsi Likelihood

Fungsi *likelihood* gabungan dari distribusi *mixture* berbeda dengan fungsi *likelihood* dari distribusi univariat biasa. Jika terdapat data pengamatan sebanyak  $n$  yang terdekomposisi ke dalam  $k$  kelompok data ( $k$  sub populasi) dimana masing-masing sub populasi memiliki distribusi, maka fungsi *likelihood* model *mixture* tersusun dari beberapa *likelihood* data di setiap sub populasi menurut distribusi masing-masing sub populasi tersebut (Mc. Lachlan dan Basford, 1988).

Berdasarkan model *mixture* pada persamaan (2.3), Mc. Lachlan dan Basford (1988) dan Astuti (2006) menyatakan bahwa fungsi *likelihood* model *mixture* seperti pada persamaan (2.12).

$$\begin{aligned}
 l_{mixture}(\boldsymbol{\theta}|\mathbf{x}) &= \sum_{j=1}^k \left( \prod_{i_j=1}^{n_j} w_j g_j(x_{i_j} | \theta_j) \right) \\
 &= \prod_{i_1=1}^{n_1} w_1 g_1(x_{i_1} | \theta_1) + \prod_{i_2=1}^{n_2} w_2 g_2(x_{i_2} | \theta_2) + \dots + \prod_{i_k=1}^{n_k} w_k g_k(x_{i_k} | \theta_k)
 \end{aligned} \tag{2.12}$$

dengan  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ ,  $n = n_1 + n_2 + \dots + n_k$  dan  $k$  adalah banyaknya komponen dalam model *mixture*.

Apabila fungsi *likelihood* gabungan ini diterapkan pada kasus model *mixture* normal maka fungsi  $g_j(x_{i_j} | \theta_j)$  dalam persamaan (2.12) diganti dengan fungsi pada persamaan (2.5) menjadi  $g_j(\mathbf{x} | \mu_j, \sigma_j^2) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp \left\{ -\frac{(\mathbf{x} - \mu_j)^2}{2\sigma_j^2} \right\}$ , dengan  $j = 1, 2, \dots, k$ , sehingga persamaan (2.12) dapat dituliskan sebagai berikut:

$$l_{mixture}(\boldsymbol{\mu}, \boldsymbol{\sigma}, \mathbf{w} | \mathbf{x}) = \sum_{j=1}^k \left( \prod_{i_j=1}^{n_j} w_j g_j(x_{i_j} | \mu_j, \sigma_j^2) \right) \\ = \prod_{i_1=1}^{n_1} w_1 g_1(x_{i_1} | \mu_1, \sigma_1^2) + \prod_{i_2=1}^{n_2} w_2 g_2(x_{i_2} | \mu_2, \sigma_2^2) + \dots + \prod_{i_k=1}^{n_k} w_k g_k(x_{i_k} | \mu_k, \sigma_k^2) \quad (2.13)$$

### 2.5.2. Distribusi Prior

Sebagaimana telah dinyatakan dalam sub bab 2.3 bahwa dalam konsep Bayesian, informasi awal (data masa lalu) yang diposisikan sebagai distribusi prior dan informasi dari sampel yang digunakan untuk membentuk fungsi *likelihood* dikombinasikan untuk membentuk distribusi posterior. Ada beberapa tipe distribusi prior yang dikenal dalam analisis Bayesian yaitu:

1. *Conjugate prior* atau *non-conjugate prior* (Box dan Tiao, 1973; Zellner, 1971; Tanner, 1996 dan Gelman dkk., 1995). Prior ini mempunyai kesamaan dan ketidaksamaan dengan pola model *likelihood* dalam data.
2. *Proper prior* atau *improper prior* (*Jeffreys prior*) yaitu prior yang pada setiap titik diberi bobot/densitas sehingga terdistribusi secara *uniform* atau tidak.
3. *Informative prior* atau *non-informative prior*, yaitu prior yang terkait dengan diketahui atau belum diketahuinya pola/frekuensi distribusi dari data.
4. *Pseudo prior* (Charlin dan Chip, 1995), yaitu prior yang nilainya disetarakan dengan hasil elaborasi dari pendapat kelompok *frequentist* (misalkan pada kasus regresi klasik dengan pendekatan OLS).



Penentuan distribusi prior untuk setiap parameter model yang akan dibentuk dilakukan dengan pendekatan struktur hirarki. Berikut ini disajikan contoh distribusi prior secara hirarki dalam pemodelan *mixture* normal sebagaimana ditunjukkan pada persamaan (2.13) (Richardson dan Green, 1997 dan Astuti, 2006).

$$\begin{aligned}\mu_j &\sim N(\xi, \kappa^{-1}), j=1,2,\dots,k \\ \sigma_j^{-2} | \beta &\sim \Gamma(\alpha, \beta), j=1,2,\dots,k \\ \beta &\sim \Gamma(g, h), \\ p(z_i = j) &= w_j \text{ dan } x_i | z_i \sim N(\mu_{z_i}, \sigma_{z_i}^2) \text{ untuk } i=1,2,\dots,n; \\ \mathbf{w} | k &\sim D(\delta_1, \delta_2, \dots, \delta_k).\end{aligned}$$

dimana prior dari  $\mu_j$  adalah  $N(\xi, \kappa^{-1})$  dengan  $\xi$  merupakan *midpoint* dari  $\mu_j$  dan  $\kappa$  adalah bilangan presisi yang nilainya sebesar  $\frac{1}{R^2}$ , dengan  $R$  adalah kisaran (*range*) dari data  $\mathbf{x}$ . Sedangkan prior dari  $\sigma_j^{-2} | \beta$  adalah  $\Gamma(\alpha, \beta)$  dengan parameter  $\alpha$  dan  $\beta$ , dimana konstanta  $\alpha$  memiliki kisaran nilai  $\alpha > 1 > g$  dan  $\beta$  adalah hiperparameter untuk  $\sigma_j^{-2}$  yang memiliki distribusi gamma dengan parameter skala  $h$  yang dapat pula diberi nilai  $\frac{1}{R^2}$  dan parameter konstanta  $g$ . Selanjutnya  $\delta_k$  adalah parameter dari distribusi dirichlet  $\mathbf{w} | k$  yang dalam hal ini berupa nilai konstanta.

Setiap data pengamatan  $x_i$  diambil secara independen dari sub populasi  $z_i$  yang tidak diketahui dengan  $p(z_i = j) = w_j, i=1,2,\dots,n; j=1,2,\dots,k$  dan  $x_i | z_i \sim N(\mu_{z_i}, \sigma_{z_i}^2)$  untuk  $i=1,2,\dots,n$ . Selanjutnya  $\mu_j$  ditentukan secara terurut sebagai berikut:  $\mu_1 < \mu_2 < \dots < \mu_k$  dengan  $j=1,2,\dots,k$ . Sedangkan untuk parameter pembobot dari setiap komponen *mixture* ditentukan sebagai  $\mathbf{w} | k \sim \text{dirichlet}(\delta_1, \delta_2, \dots, \delta_k)$ .

### 2.5.3. Distribusi Bersyarat Penuh

Distribusi bersyarat penuh dibentuk untuk memudahkan estimasi parameter dalam model *mixture* yang seringkali cukup kompleks dan bahkan untuk beberapa kasus cukup rumit sehingga parameter-parameternya sulit untuk diestimasi. Menurut Gilks (1995), distribusi bersyarat penuh dari suatu parameter  $\Lambda = (\mathbf{z}, \boldsymbol{\theta}, \mathbf{w}, k)$  yang memiliki distribusi  $p(\Lambda)$  yang stasioner dan dibentuk dengan cara membuat partisi dari  $\Lambda$  sebagaimana yang ditunjukkan dalam persamaan (2.14).

$$\Lambda = (\Lambda_k, \Lambda_{-k}), \quad (2.14)$$

dimana  $\Lambda_k$  menotasikan vektor parameter  $\Lambda$  komponen ke- $k$  yang akan diestimasi dalam model *mixture* dan  $\Lambda_{-k}$  menotasikan komplemen dari  $\Lambda_k$ , yaitu vektor parameter  $\Lambda$  tanpa memuat parameter dalam komponen ke- $k$ . Sebagai contoh, andaikan parameter  $\boldsymbol{\theta}$  diestimasi maka  $\Lambda_k = \boldsymbol{\theta}$  dan  $\Lambda_{-k} = (\mathbf{z}, \mathbf{w}, k)$ .

Selanjutnya Gilks (1995), Suharjo (2003), dan Wati (2006) menjelaskan bahwa pembentukan distribusi bersyarat penuh dapat diperoleh dari bentuk distribusi gabungan semua parameter dalam model dengan syarat parameter yang lain dianggap konstan. Bentuk distribusi bersyarat penuh ditunjukkan dalam persamaan (2.15).

$$p(\Lambda_k | \Lambda_{-k}) = \frac{p(\Lambda_k, \Lambda_{-k})}{\int p(\Lambda_k, \Lambda_{-k}) d\Lambda_k}. \quad (2.15)$$

Untuk menentukan distribusi bersyarat penuh dari model *mixture* normal adalah dengan manipulasi dari bentuk distribusi gabungan semua parameter dalam model *mixture* normal, sebagaimana yang telah disajikan dalam persamaan (2.11). Distribusi bersyarat penuh untuk masing-masing parameter dalam model *mixture* normal dijabarkan sebagai berikut (Gilks, 1995 dan Astuti, 2006):

### a. Estimasi Parameter Lokasi ( $\mu_j$ )

Distribusi prior dari parameter  $\mu_j$  adalah  $normal(\xi, K^{-1})$  dan diketahui konsep teorema Bayes sebagai berikut berikut:

$$p(\mu_j | \dots) \propto \frac{p(\mu_j, \sigma_j^{-2} | \mathbf{x})}{p(\sigma_j^{-2} | \mathbf{x})} = \frac{p(\mathbf{x}, \mu_j, \sigma_j^{-2})}{p(\mathbf{x}, \sigma_j^{-2})} \propto p(\mathbf{x}, \mu_j, \sigma_j^{-2}).$$

Distribusi bersyarat penuh dari parameter lokasi  $\mu_j$  diperoleh dari rumusan berikut ini:

$$p(\mu_j | \sigma^2, \mathbf{z}, \mathbf{w}, k) \propto l(\mathbf{x} | \sigma^2, \mathbf{z}, \mathbf{w}, k) p(\mu_j | \mathbf{z}, \mathbf{w}, k),$$

Penjabaran secara matematis dari persamaan distribusi bersyarat penuh untuk parameter lokasi sebagaimana ditunjukkan dalam persamaan (2.16) (Gilks, 1995; Astuti, 2006).

$$\begin{aligned} p(\mu_j | \dots) &\propto \exp \left\{ -\frac{\sigma_j^{-2}}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 \right\} \times \exp \left\{ -\frac{K}{2} (\mu_j - \xi)^2 \right\} \\ &= \exp \left\{ -\frac{1}{2} \left( \left( \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 \right) + K (\mu_j - \xi)^2 \right) \right\} \\ &= \exp \left\{ -\frac{1}{2} \left( \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} x_i^2 - 2\sigma_j^{-2} \mu_j \sum_{i=1; z_i=j}^{n_j} x_i + \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} \mu_j^2 + K \mu_j^2 - 2K\xi \mu_j + K\xi^2 \right) \right\} \\ &= \exp \left\{ -\frac{1}{2} \left( \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} x_i^2 - 2\mu_j \left( \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} x_i + K\xi \right) + \mu_j^2 (\sigma_j^{-2} n_j + K) + K\xi^2 \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (K + \sigma_j^{-2} n_j) \left( \mu_j^2 - \frac{2\mu_j \left( \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} x_i + K\xi \right)}{K + \sigma_j^{-2} n_j} + \frac{\sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} x_i^2 + K\xi^2}{K + \sigma_j^{-2} n_j} \right) \right\} \\ &\propto \exp \left\{ -\frac{1}{2} (K + \sigma_j^{-2} n_j) \left( \mu_j - \frac{\left( \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} x_i + K\xi \right)}{K + \sigma_j^{-2} n_j} \right)^2 \right\}. \end{aligned} \tag{2.16}$$

Bentuk terakhir dari persamaan (2.16) adalah bentuk distribusi normal, sehingga distribusi bersyarat penuh dari parameter lokasi  $\mu_j$  ditunjukkan dalam persamaan (2.17).

$$p(\mu_j | \sigma^2, \mathbf{z}, \mathbf{w}, k) \sim N \left( \frac{\left( \sigma_j^{-2} \sum_{i=1; z_i=j}^{n_j} x_i + K\xi \right)}{K + \sigma_j^{-2} n_j}, \left( K + \sigma_j^{-2} n_j \right)^{-1} \right) \quad (2.17)$$

dengan  $j = 1, 2, \dots, k$ .

### b. Estimasi Parameter Varians ( $\sigma_j^{-2}$ )

Distribusi prior dari parameter  $\sigma_j^{-2}$  adalah normal, sehingga bentuk distribusi bersyarat penuh dari parameter  $\sigma_j^{-2}$  adalah:

$$p(\sigma_j^{-2} | \mu_j, \mathbf{z}, \mathbf{w}, k) \propto l(\mathbf{x} | \sigma^2, \mathbf{z}, \mathbf{w}, k) p(\sigma_j^{-2} | \mathbf{z}, \mathbf{w}, k).$$

Penjabaran secara matematis untuk distribusi bersyarat penuh dari parameter varians  $\sigma_j^{-2}$  dapat ditulis sebagaimana pada persamaan (2.18).

$$\begin{aligned} p(\sigma_j^{-2} | \dots) &\propto \left( \sigma_j^{-2} \right)^{\frac{n_j}{2}} \exp \left\{ -\frac{\sigma_j^{-2}}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 \right\} \frac{\beta^\alpha \left( \sigma_j^{-2} \right)^{\alpha-1} \exp(-\beta \sigma_j^{-2})}{\Gamma(\alpha)} \\ &= \frac{\beta^\alpha \left( \sigma_j^{-2} \right)^{\frac{n_j}{2} + \alpha - 1} \exp \left\{ -\frac{\sigma_j^{-2}}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 - \beta \sigma_j^{-2} \right\}}{\Gamma(\alpha)} \\ &= \frac{\beta^\alpha \left( \sigma_j^{-2} \right)^{\frac{n_j}{2} + \alpha - 1} \exp \left\{ -\sigma_j^{-2} \left( \beta + \frac{1}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 \right) \right\}}{\Gamma(\alpha)} \quad (2.18) \\ &\propto \frac{\left( \beta + \frac{1}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 \right)^{-\frac{n_j}{2} - \alpha} \left( \sigma_j^{-2} \right)^{\frac{n_j}{2} + \alpha - 1} \exp \left\{ -\sigma_j^{-2} \left( \beta + \frac{1}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 \right) \right\}}{\Gamma \left( \alpha + \frac{1}{2} n_j \right)} \\ &\propto \Gamma \left( \alpha + \frac{1}{2} n_j, \beta + \frac{1}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2 \right). \end{aligned}$$

Sehingga distribusi bersyarat penuh untuk parameter varians  $\sigma_j^{-2}$  dapat ditulis dalam persamaan (2.19).

$$p(\sigma_j^{-2} | \mu_j, \mathbf{w}, k) \sim \Gamma\left(\alpha + \frac{1}{2}n_j, \beta + \frac{1}{2} \sum_{i=1; z_i=j}^{n_j} (x_i - \mu_j)^2\right). \quad (2.19)$$

**c. Estimasi Parameter Alokasi (  $z_i$  )**

Distribusi prior dari parameter alokasi  $z_i$  adalah  $Normal(\mu_{z_i}, \sigma_{z_i}^2)$ , sehingga bentuk distribusi bersyarat penuh dari parameter  $z_i$  adalah:

$$p(z_i = j | \mathbf{w}, k) \propto w_j \frac{1}{\sqrt{2\pi}} \sigma_j^{-2} \exp\left\{-\frac{\sigma_j^{-2}}{2} (x_i - \mu_j)^2\right\}.$$

Jika persamaan ini dijabarkan secara matematis, maka akan diperoleh informasi bahwa distribusi bersyarat penuh dari parameter  $z_i$  dapat ditulis dalam persamaan (2.20).

$$p(z_i | \mathbf{w}, k) \propto \frac{w_j}{\sigma_j} \sigma_j^{-2} \exp\left\{-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}\right\}. \quad (2.20)$$

**d. Estimasi Parameter Pembobot (  $w_j$  )**

Distribusi prior parameter pembobot  $w_j$  adalah  $dirichlet(\delta_1, \delta_2, \dots, \delta_k)$ , sehingga bentuk distribusi bersyarat penuh dari parameter pembobot  $w_j$  adalah:

$$p(\mathbf{w} | k) \propto \prod_{i=1}^n \left[ \frac{\Gamma\left(\sum_{j=1}^k \delta_j\right)}{\prod_{j=1}^k \Gamma(\delta_j)} \prod_{j=1}^k w_j^{\delta_j-1} \right].$$

Penjabaran persamaan distribusi bersyarat penuh dari parameter pembobot  $w_j$  dapat ditulis dalam persamaan (2.21).

$$\begin{aligned}
p(\mathbf{w} | \dots) &= \prod_{i=1}^n \left[ \frac{\Gamma\left(\sum_{j=1}^k \delta\right)}{\prod_{j=1}^k \Gamma(\delta)} \prod_{j=1}^k w_j^{\delta-1} \right] \propto \prod_{i=1}^n \prod_{j=1}^k w_j^{\delta-1} = \prod_{j=1}^k \prod_{i=1; z_i=j}^{n_j} w_j^{\delta-1}; n = n_1 + \dots + n_k \\
&= \prod_{j=1}^k w_j^{\delta-1 \times n_j} = \prod_{j=1}^k w_j^{\delta+n_j-1} \propto \frac{\Gamma\left(\sum_{j=1}^k (\delta+n_j)\right)}{\prod_{j=1}^k \Gamma(\delta+n_j)} \prod_{j=1}^k w_j^{\delta+n_j-1}.
\end{aligned} \tag{2.21}$$

Bentuk terakhir dari persamaan (2.21) adalah bentuk dari distribusi *dirichlet*, yaitu  $D[\delta+n_1, \delta+n_2, \dots, \delta+n_k]$ . Sehingga distribusi bersyarat penuh untuk parameter pembobot  $w_j$  sebagaimana ditulis dalam persamaan (2.21).

$$p(\mathbf{w} | k) \sim D(\delta_{j_1} + n_1, \delta_{j_2} + n_2, \dots, \delta_{j_k} + n_k). \tag{2.22}$$

#### e. Estimasi *Hyperparameter* ( $\beta$ )

Parameter  $\beta$  merupakan hiperparameter dari  $\sigma_j^{-2}$  yang memiliki distribusi prior  $\text{gamma}(g, h)$ , sehingga bentuk distribusi bersyarat penuh dari hiperparameter  $\beta$  adalah:

$$p(\beta | \sigma_j^{-2}) \propto p(\sigma_j^{-2} | \alpha, \beta) p(\beta | g, h).$$

Penjabaran secara matematis bentuk distribusi bersyarat penuh dari *hyperparameter*  $\beta$  sebagaimana ditunjukkan dalam persamaan (2.23).

$$\begin{aligned}
p(\beta | \dots) &\propto p(\sigma_j^{-2} | \alpha, \beta) p(\beta | g, h) \\
&= \frac{\beta^{k\alpha} (\sigma_j^{-2})^{\alpha-1} \exp\left\{-\sum_{j=1}^k \beta \sigma_j^{-2}\right\}}{\Gamma(k\alpha)} \times \frac{h^g \beta^{g-1} \exp(-\beta h)}{\Gamma(g)} \\
&\propto \frac{\beta^{k\alpha+g-1} \exp\left\{-\beta\left(h + \sum_{j=1}^k \sigma_j^{-2}\right)\right\} (\sigma_j^{-2})^{k\alpha-1} h^g}{\Gamma(g+k\alpha)} \\
&\propto \frac{\beta^{k\alpha+g-1} \exp\left\{-\beta\left(h + \sum_{j=1}^k \sigma_j^{-2}\right)\right\} \left(h + \sum_{j=1}^k \sigma_j^{-2}\right)^{(g+k\alpha)}}{\Gamma(g+k\alpha)} \\
&= \Gamma\left(g+k\alpha-1, h + \sum_{j=1}^k \sigma_j^{-2}\right).
\end{aligned} \tag{2.23}$$

Sehingga distribusi bersyarat penuh dari hiperparameter  $\beta$  dapat ditulis dalam persamaan (2.24).

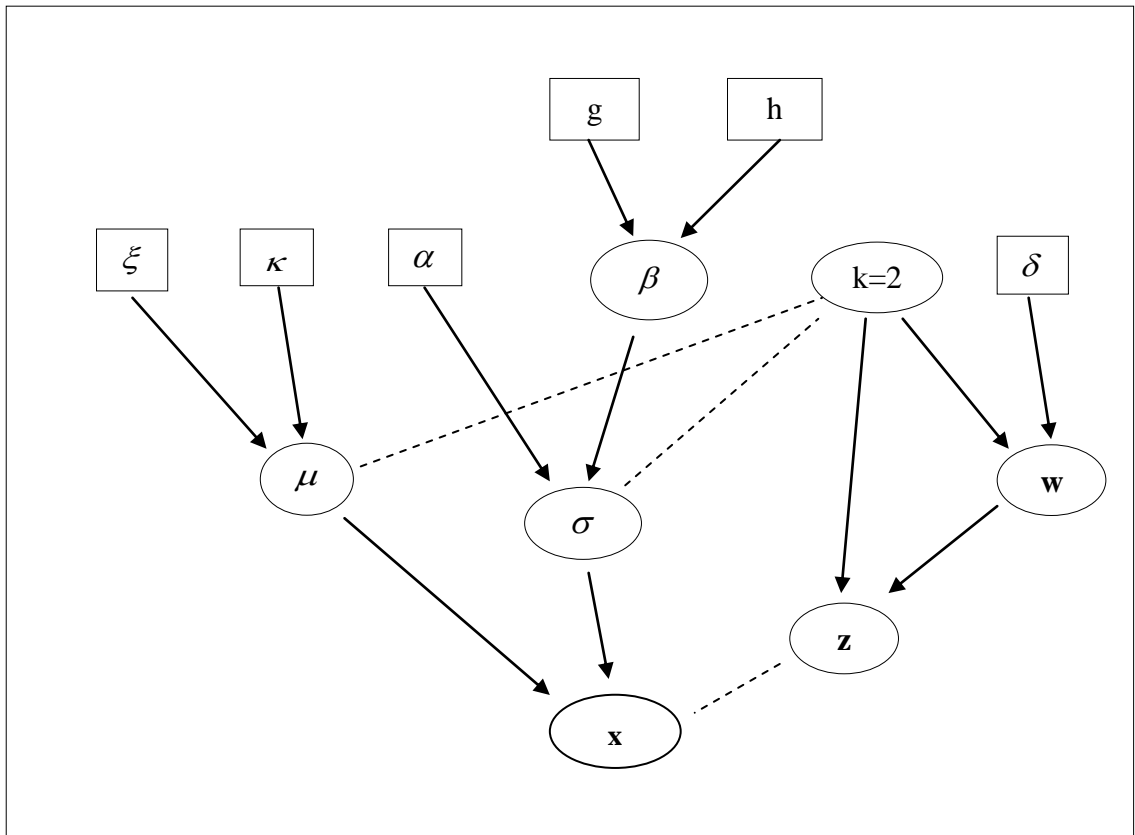
$$p(\beta | \sigma_j^{-2}) \sim \Gamma\left(g+k\alpha-1, h + \sum_{j=1}^k \sigma_j^{-2}\right). \tag{2.24}$$

#### 2.5.4. Directed Acyclic Graph (DAG) Model Mixture

*Directed Acyclic Graph* (DAG) merupakan suatu *graph* yang mampu menggambarkan struktur hirarki dari suatu model (Ntzoufras, 2009). Andaikan DAG akan digunakan untuk model *mixture*, maka struktur hirarki dari model *mixture* sebagaimana pada persamaan (2.3) dikatakan sebagai model hirarki Bayesian karena struktur model priornya mempunyai susunan yang bertingkat dan sistematis, sehingga dapat digambarkan ke dalam bentuk DAG (Iriawan, 2003b; Astuti, 2006 dan Dinar, 2009).

DAG dapat disusun dalam dua *node* yang berbeda yaitu kotak dan ellips (Iriawan, 2003b). *Node* kotak menunjukkan bahwa parameter bernilai konstan, sedangkan *node ellips* menunjukkan bahwa parameter bisa berubah secara *stochastic* ataupun *logical*. Keduanya dibedakan oleh *directed graph* inputnya yang berupa garis tunggal atau garis putus-putus, dimana garis tunggal untuk menyatakan hubungan parameter yang bersifat *stochastic* dan garis putus-putus untuk menyatakan hubungan parameter yang bersifat *logical*. Berikut ini akan

diberikan contoh suatu struktur hirarki model *mixture* dengan jumlah komponen penyusun sebanyak dua ( $k=2$ ) dimana masing-masing komponen berdistribusi normal (Stephens, 1997; Iriawan, 2001 dan Dinar, 2009) sebagaimana ditunjukkan dalam Gambar 2.1. Nilai  $k=2$  merupakan banyaknya komponen penyusun model *mixture* normal,  $\mathbf{w}$  menotasikan pembobot dari setiap komponen penyusun model *mixture* normal,  $\mathbf{z}$  menotasikan alokasi setiap data pada setiap sub populasi,  $\mu$  dan  $\sigma$  menotasikan parameter komponen penyusun model *mixture* yang berdistribusi normal, sedangkan  $\mathbf{x}$  adalah data pengamatan.



**Gambar 2.1.** Struktur Hirarki dari Prior Model *Mixture* Normal dengan Dua Komponen Penyusun dalam DAG

#### 2.5.5. Metode Markov Chain Monte Carlo (MCMC)

Untuk mendapatkan distribusi posterior dalam analisis Bayesian seringkali sangat rumit dilakukan dan memerlukan proses integrasi yang sulit dalam menentukan marginal posterior suatu parameter model sehingga dibutuhkan suatu metode pendekatan numerik *Markov Chain Monte Carlo* (MCMC) (Iriawan, 2000). Melalui pendekatan MCMC akan memudahkan pemodelan yang cukup



komplek sehingga metode ini dianggap sebagai suatu terobosan dalam penggunaan analisis Bayesian (Hastings, 1970; Charlin dan Chip, 1995). Metode MCMC merupakan metode simulasi yang menggabungkan antara *Monte Carlo* dengan sifat *Markov Chain* untuk mendapatkan data sampel berdasarkan skenario sampling tertentu. Menurut Iriawan (2000), Rantai Markov pada ruang parameter  $\Omega$  didefinisikan sebagai suatu deret parameter random  $\{\theta_j\}_{j>0}$  dengan  $j=1,2,\dots,k$ , dimana nilai untuk masing-masing parameter random berada di dalam ruang parameter  $\Omega$  dan distribusi dari  $\theta_j$  yang telah diberikan semua nilai sebelumnya dari proses, yaitu  $\theta_1, \theta_2, \dots, \theta_{j-1}$ , hanya tergantung pada  $\theta_{j-1}$ . Secara matematis, definisi dari rantai Markov dapat dijabarkan dalam Definisi 2.1 (Iriawan, 2000).

**Definisi 2.1.** Definisi Rantai Markov

*Misalkan  $\{\theta_j\}_{j>0}$  merupakan deret dari suatu parameter random. Deret ini dikatakan sebagai suatu rantai Markov jika diberikan suatu nilai untuk  $\theta_j$  sedemikian hingga distribusi bersyarat  $\theta_j$  dengan  $\theta_1, \theta_2, \dots, \theta_{j-1}$  diketahui hanya akan bergantung pada nilai  $\theta_{j-1}$  saja atau dapat dituliskan sebagai berikut:*

$P(\theta_i \in \Omega | \theta_{j-1} \in \Omega_{j-1}, \theta_{j-2} \in \Omega_{j-2}, \dots, \theta_0 \in \Omega_0) = P(\theta_j \in \Omega | \theta_{j-1} \in \Omega_{j-1})$  untuk semua  $\Omega_0, \Omega_1, \dots, \Omega_{j-1}; \Omega_k \subset \Omega$ , dengan  $\Omega_k$  merupakan himpunan bagian dari  $\Omega$  dimana  $\Omega$  adalah ruang parameter.

Penggunaan metode MCMC pada analisis Bayesian membutuhkan skenario sampling yang tepat untuk mendapatkan sampel dari suatu distribusi tertentu yang diketahui baik secara langsung maupun tidak langsung. Jika suatu variabel random diketahui berdistribusi  $f(\mathbf{x}|\boldsymbol{\theta}, \mathbf{w})$  yang diketahui maka pembangkitan data dapat dilakukan secara langsung. Sebaliknya jika  $f(\mathbf{x}|\boldsymbol{\theta}, \mathbf{w})$  tidak berbentuk distribusi yang umum diketahui atau distribusi yang kompleks, maka proses pembangkitan data tidak dapat dilakukan secara langsung (Iriawan, 2000).

Dalam metode MCMC, salah satu pendekatan yang umum digunakan untuk  $f(\mathbf{x}|\boldsymbol{\theta}, \mathbf{w})$  yang berbentuk kompleks atau tidak umum adalah metode *Gibbs Sampler*, dimana metode pendekatan ini merupakan suatu teknik simulasi untuk membangkitkan variabel random dari suatu distribusi tertentu secara tidak langsung tanpa harus menghitung densitasnya dan merupakan *generator* yang sangat efisien. Metode *Gibbs Sampler* ini didasarkan hanya pada sifat-sifat dasar dari MCMC sehingga perhitungan yang sulit dapat dihindari (Iriawan, 2000; Casella dan George, 1992 dan Astuti, 2006). Cara kerja dari metode *Gibbs Sampler* adalah sebagai berikut: (Casella dan George, 1992; Gamerman, 1997; Walsh, 2002 dan Astuti, 2006).

Misalkan  $\mathbf{x}$  adalah sampel random yang memiliki distribusi stasioner  $p(\boldsymbol{\theta}|\mathbf{x})$  dimana  $\boldsymbol{\theta}$  adalah parameter, kemudian  $\mathbf{x}$  dan  $\boldsymbol{\theta}$  dapat dipartisi menjadi komponen-komponen  $(x_1, x_2, \dots, x_r) \in S_1 \times S_2 \times \dots \times S_r$  dan  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_r)$  dengan  $S_r$  merupakan himpunan bagian ke- $r$  dari *state space*  $\mathbf{S}$ . Proses sampling secara langsung dari  $p(\boldsymbol{\theta}|\mathbf{x}) = p(\theta_1, \theta_2, \dots, \theta_r | \mathbf{x})$  tidak dapat dilakukan karena distribusi tersebut sangat kompleks dimana kekomplekkanannya disebabkan oleh banyaknya parameter dalam model. Untuk mempermudah estimasi terhadap setiap parameter dapat dilakukan dengan cara membangkitkan sampel dari distribusi bersyarat penuh setiap parameter  $(\theta_j; j = 1, 2, \dots, r)$  terhadap komplemennya  $(\theta_{-j})$  dan data  $\mathbf{x}$ , yaitu  $p(\theta_1 | x, \theta_2, \dots, \theta_r), \dots, p(\theta_r | x, \theta_1, \dots, \theta_{r-1})$ . Data yang dibangkitkan dengan sistem ini akan menghasilkan pola data yang stasioner dan konvergen (Stephens, 1997).

Langkah-langkah dari Algoritma MCMC dengan pendekatan metode *Gibbs Sampler* dapat dijabarkan dalam Algoritma 2.1 (Gamerman, 1997; Stephens, 1997; Walsh, 2002 dan Wati, 2006).

**Algoritma 2.1.** Algoritma MCMC Pendekatan Metode *Gibbs Sampler*

Apabila parameter model dinotasikan  $\boldsymbol{\theta}$ , maka

Langkah 1. Berikan nilai awal untuk  $\boldsymbol{\theta}^{(k)}$  pada  $k = 0$  sehingga

$$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \dots, \theta_r^{(0)})$$

Langkah 2. Melakukan proses sampling untuk mendapatkan nilai dari  $\theta_j$  dari distribusi bersyarat dengan cara membangkitkan  $\theta^{(k+1)}$  dalam  $r$  langkah sebagai berikut:

1. Sampling  $\theta_1^{(k+1)}$  dari  $p(\theta_1 | X, \theta_2^{(k)}, \dots, \theta_r^{(k)})$
  2. Sampling  $\theta_2^{(k+1)}$  dari  $p(\theta_2 | X, \theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_r^{(k)})$
  - $\vdots$
  - r. Sampling  $\theta_r^{(k+1)}$  dari  $p(\theta_r | X, \theta_1^{(k+1)}, \theta_2^{(k+1)}, \dots, \theta_{r-1}^{(k+1)})$
- $k \leftarrow k + 1$

Langkah 3. Melakukan iterasi sampai tercapai konvergen pada Langkah 2 sebanyak  $M$  kali dengan  $M \rightarrow \infty$ .

Beberapa asumsi yang harus dipenuhi dalam menggunakan metode MCMC pada analisis Bayesian agar diperoleh hasil estimasi parameter-parameter model yang tepat (Casella dan George, 1992; Iriawan, 2000; Walsh, 2002), yaitu:

#### 1. *Independence*

Sifat *independence* dari setiap parameter model yang diestimasi akan terpenuhi jika plot fungsi autokorelasi dari semua parameter model yang diestimasi memiliki nilai nyata berbeda dengan nol untuk lag-0 dan lag-lag selain lag-0 mempunyai nilai tidak nyata berbeda dengan nol.

#### 2. *Irreducible, Aperiodic dan Recurrent*

Sifat ini terkait dengan konvergensi dari setiap parameter model yang diestimasi. Hal ini dapat ditunjukkan dari plot distribusi Kernel, *history*, dan *quantiles* dari setiap parameter yang diestimasi telah dinyatakan konvergen jika telah memenuhi sifat *irreducible*, *aperodic* dan *recurrent*. Sifat ini ditunjukkan oleh setiap parameter yang diestimasi memiliki plot distribusi Kernel yang lebih mulus (*smooth*), plot *history* yang relatif homogen dari satu iterasi ke iterasi yang lain dan plot *quantiles* berada pada batas-batas *quantiles* yang ditetapkan untuk semua iterasi yang dilakukan.

Apabila sifat-sifat *independence*, *irreducible*, *aperodic* dan *recurrent* dari semua parameter model yang diestimasi telah terpenuhi maka model yang dibangun dengan MCMC pada analisis Bayesian sudah tepat. Jika asumsi-asumsi

ini tidak terpenuhi maka perlu dilakukan perbaikan-perbaikan model dengan mencoba beberapa cara perbaikan, seperti menambah jumlah iterasi, merubah nilai *thin*, merubah distribusi dari data maupun distribusi priornya.

### 2.5.6. *Metropolis-Hastings*

Metode penyusunan rantai Markov dengan Algoritma Metropolis-Hastings dijelaskan oleh Gamerman (1997) dan Wati (2006) sebagai berikut: misalkan diketahui suatu parameter  $\theta_1$  dan suatu parameter  $\theta_2$ , kemudian terdapat probabilitas transisi Kernel  $P(\theta_1, \theta_2)$  dengan transisi Kernel dari parameter  $\theta_1$  adalah  $\pi(\theta_1)$  dan transisi Kernel dari parameter  $\theta_2$  adalah  $\pi(\theta_2)$  yang memenuhi persamaan (2.25).

$$\pi(\theta_1)P(\theta_1, \theta_2) = \pi(\theta_2)P(\theta_2, \theta_1), \forall (\theta_1, \theta_2) \quad (2.25)$$

Selanjutnya probabilitas transisi Kernel  $P(\theta_1, \theta_2)$  yang terdiri atas dua komponen adalah  $q(\theta_1, \theta_2)$  dan  $\alpha(\theta_1, \theta_2)$  sehingga  $P(\theta_1, \theta_2) = q(\theta_1, \theta_2) \alpha(\theta_1, \theta_2)$  untuk  $\theta_1 \neq \theta_2$ , dimana  $q(\theta_1, \theta_2)$  menotasikan densitas dari  $\theta_1$  ke *state* parameter  $\theta_2$  yang akan diajukan untuk mendapatkan rantai markov berikutnya dan  $\alpha(\theta_1, \theta_2)$  menotasikan probabilitas penerimaan untuk pergerakan dari  $\theta_1$  ke  $\theta_2$  yang dinyatakan dalam persamaan (2.25).

$$\alpha(\theta_1, \theta_2) = \min \left[ 1, \frac{\pi(\theta_2)q(\theta_2, \theta_1)}{\pi(\theta_1)q(\theta_1, \theta_2)} \right], \quad (2.26)$$

dengan  $q(\theta_2, \theta_1)$  menotasikan densitas dari parameter  $\theta_2$  untuk kembali ke *state* parameter  $\theta_1$ . Jika suatu titik kandidat dari parameter  $\theta_2$  diterima sebagai sampel, maka *state* berikutnya menjadi  $\theta_2$ , sedangkan jika suatu titik kandidat dari parameter  $\theta_2$  ditolak sebagai sampel, maka rantai Markov tidak berubah, sehingga  $\theta_2 = \theta_1$ . Dengan demikian, probabilitas transisi Kernel  $P(\theta_1, \theta_2)$  tidak selalu menggambarkan semua transisi dari parameter  $\theta_1$  ke parameter  $\theta_2$ . Hal ini karena ada kemungkinan penolakan transisi dari parameter  $\theta_1$  ke parameter  $\theta_2$ . Oleh

karena itu, jika terjadi penolakan transisi dari parameter  $\theta_1$  ke parameter  $\theta_2$ , maka rantai Markov tidak akan berubah, sehingga probabilitas terjadinya penolakan transisi dari parameter  $\theta_1$  ke parameter  $\theta_2$  yang disimbolkan dengan  $P(\theta_1, \theta_1) = r(\theta_1)$  dinyatakan dalam persamaan (2.27).

$$P(\theta_1, \theta_1) = r(\theta_1) = 1 - \int_{-\infty}^{\infty} q(\theta_1, \theta_2) \alpha(\theta_1, \theta_2) d\theta_2. \quad (2.27)$$

Langkah-langkah dari Algoritma *Metropolis-Hastings* dapat dijelaskan dalam Algoritma 2.2 (Gamerman, 1997; Walsh, 2002 dan Wati, 2006).

**Algoritma 2.2.** Algoritma *Metropolis-Hastings*

Apabila suatu parameter  $\theta_1$  parameter  $\theta_2$  diketahui, maka

Langkah 1. Inisialisasi  $\theta_0$ , dimana  $t = 0$

Langkah 2. Bangkitkan sampel dari parameter  $\theta_2$  dari  $q(\bullet | \theta_t)$

Langkah 3. Bangkitkan variabel random  $\mathbf{u} \sim U(0,1)$

IF  $\mathbf{u} \leq \alpha(\theta_t, \theta_2)$  set  $\theta_{t+1} = \theta_2$

ELSE berikan  $\theta_{t+1} = \theta_1$

$t \leftarrow t + 1$

Langkah 4. Ulangi Langkah 2 dan Langkah 3 sampai tercapai konvergen sebanyak  $t$  kali dengan  $t \rightarrow \infty$ .

**2.5.7. Reversible Jump**

Dalam penggunaan Algoritma MCMC, proses iterasi Bayesian hanya terbatas pada kasus dimensi vektor dari parameter model yang bersifat tetap (diketahui). Dalam berbagai kasus ditemukan bahwa dimensi vektor parameter model bersifat tidak tetap terutama yang disebabkan oleh model *mixture* dengan jumlah komponen *mixture* yang tidak/belum diketahui banyaknya. Green (1995) telah mengembangkan Algoritma *Metropolis-Hastings* untuk mengatasi permasalahan dimensi vektor parameter model bersifat tidak tetap yang disebut dengan *Reversible Jump Markov Chain Monte Carlo* (RJMC MC). RJMC MC pada pemodelan *mixture* dapat digunakan untuk menentukan jumlah komponen

*mixture* yang tidak diketahui secara pasti. Dalam hal ini berarti bahwa dimensi vektor dari parameter model *mixture* bersifat tidak tetap (tidak diketahui besarnya). Richardson dan Green (1997) serta Wati (2006) memanfaatkan konsep RJMCMC untuk mendapatkan model *mixture* yang tepat dimana banyaknya komponen penyusun *mixture* tidak diketahui.

Dalam aplikasinya, Algoritma RJMCMC menggunakan konsep *birth/death* dan *split/merge*. Menurut Richardson dan Green (1997) serta Wati (2006), ada enam langkah dalam Algoritma RJMCMC yaitu:

- Langkah 1. *Updating*  $\mathbf{w}$
- Langkah 2. *Updating*  $\boldsymbol{\theta}$
- Langkah 3. *Updating*  $\mathbf{z}$
- Langkah 4. *Updating* Hiperparameter  $\beta$
- Langkah 5. *Split/Merge* komponen-komponen
- Langkah 6. *Birth/Death* dari komponen kosong

Langkah 1 sampai dengan Langkah 4 tidak akan mengubah dimensi vektor parameter model yang terdiri atas  $(\beta, \boldsymbol{\theta}, k, \mathbf{w}, \mathbf{z})$ , sedangkan Langkah 5 dan Langkah 6 mengubah banyaknya komponen penyusun *mixture* ke- $k$  dengan cara satu per satu. Langkah *split/merge* merupakan pilihan random antara *split* ( $j_1 \rightarrow j_1 + 1$ ) atau *merge* ( $j_1 + 1 \rightarrow j_1$ ) dengan  $j_1 = 1, 2, \dots, k$ . Adapun langkah-langkah proses *merge*, *split*, *birth*, *death* dijabarkan sebagai berikut:

Langkah-langkah dalam proses *merge* adalah:

Anggap ada dua group yang akan bergabung, namakan group  $j_1$  dan group  $j_2$ , maka langkah-langkah proses *merge* adalah sebagai berikut:

- Langkah 1. Gabungkan pasangan yang berseberangan  $(j_1, j_2) \rightarrow j^*$
- Langkah 2. Realokasi  $\theta_j$  dengan  $z_j \in \{j_1, j_2\}$
- Langkah 3. Tentukan nilai untuk  $(w_{j^*}, \mu_{j^*}, \sigma_{j^*})$  dengan persamaan berikut:

$$w_{j^*} = w_{j_1} + w_{j_2}$$

$$w_{j^*} \mu_{j^*} = w_{j_1} \mu_{j_1} + w_{j_2} \mu_{j_2}$$

$$w_{j^*}(\mu_{j^*}^2 + \sigma_{j^*}^2) = w_{j_1}(\mu_{j_1}^2 + \sigma_{j_1}^2) + w_{j_2}(\mu_{j_2}^2 + \sigma_{j_2}^2)$$

Langkah-langkah proses *split* terdiri atas:

Anggap ada dua group yang akan memisah, namakan group  $j_1$  dan group  $j_2$ , maka langkah-langkah proses *split* adalah sebagai berikut:

Langkah 1. Komponen terpilih secara random dipisah sehingga  $j^* \rightarrow j_1 + j_2$

Langkah 2. Membangun vektor random  $\mathbf{u} = (u_1, u_2, u_3)$  dengan  $0 < u_c < 1, c = 1, 2, 3$  untuk membentuk parameter baru yang sesuai dengan persamaan berikut:

$$w_{j_1} = w_{j^*} u_1$$

$$w_{j_2} = w_{j^*} (1 - u_1)$$

$$\mu_{j_1} = \mu_{j^*} - u_2 \sigma_{j^*} \sqrt{w_{j_2} / w_{j_1}}$$

$$\mu_{j_2} = \mu_{j^*} - u_2 \sigma_{j^*} \sqrt{w_{j_1} / w_{j_2}}$$

$$\sigma_{j_1}^2 = u_3 (1 - u_2^2) \sigma_{j^*}^2 w_{j^*} / w_{j_1}$$

$$\sigma_{j_2}^2 = (1 - u_3) (1 - u_2^2) \sigma_{j^*}^2 w_{j^*} / w_{j_2}$$

Langkah 3. Ulangi Langkah 2 sampai diperoleh kondisi  $\mu_{j_1} < \mu_{j_2}$

Adapun langkah-langkah proses *birth/death* merupakan pilihan random antara *birth* ( $j_1 \rightarrow j_1 + 1$ ) atau *death* ( $j_1 + 1 \rightarrow j_1$ ). Secara lengkap, langkah-langkah dari proses *birth* dan *death* adalah sebagai berikut:

Langkah-langkah dalam proses *birth* adalah:

Langkah 1. Membentuk parameter ( $w_{j^*}, \mu_{j^*}, \sigma_{j^*}$ ) sebagaimana rumusan pada proses *merge* untuk komponen baru dengan prior  $w_{j^*} \sim \text{Beta}(1, k)$  dan  $w_j + w_{j^*} = 1$

Langkah 2. Melakukan penyekalaan kembali untuk parameter pembobot  $\mathbf{w}$  sehingga diperoleh  $\sum_{j=1}^k w_j = 1$ .

Langkah-langkah dalam proses *death* terdiri atas:

Langkah 1. Menghapus komponen kosong dari proses yang dilakukan

Langkah 2. Melakukan penyekalaan kembali dari pembobot  $\mathbf{w}$  yang tersisa.

Penerapan dari konsep RJMCMC ini telah banyak dilakukan oleh banyak peneliti, salah satunya oleh Feng (2013) untuk menentukan jumlah komponen *mixture* dalam data dengan banyaknya komponen *mixture* tidak/belum diketahui. Sebagai contoh penerapan konsep RJMCMC dengan proses *merge*, *split*, *birth*, *death* yang dibangun oleh Feng (2013), disajikan *code* program R secara lengkap di Lampiran 7.

## 2.6. Model Bayesian Unimodal Normal (BUM-Normal)

Model Bayesian Unimodal (BUM) merupakan salah satu pemodelan dengan pendekatan Bayesian untuk data yang memiliki distribusi unimodal (bukan *mixture*). Prinsip dasar untuk membentuk model BUM sama dengan prinsip dasar membentuk model BMM, perbedaannya adalah bentuk fungsi *likelihood* dan prior dari data memiliki distribusi unimodal untuk membentuk distribusi posterior dari parameter model.

Untuk membentuk model BUM-normal, fungsi *likelihood* dari data harus memiliki distribusi normal unimodal sebagaimana persamaan (2.28).

$$f(\mathbf{x} | \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2\sigma^2}\right\} \quad (2.28)$$

dimana:

$f(\mathbf{x} | \mu, \sigma^2)$  = fungsi *likelihood* normal data  $\mathbf{x}$  dengan parameter rata-rata  $\mu$  dan varians  $\sigma^2$ , dengan  $-\infty < \mu < \infty$  dan  $\sigma^2 > 0$ .

Langkah selanjutnya menentukan prior yang paling tepat dari parameter model sehingga terbentuk distribusi posterior model BUM-normal yang paling sesuai untuk data. Distribusi posterior ini merupakan distribusi posterior gabungan dari semua parameter model BUM-normal yang dibentuk dengan menerapkan Algoritma MCMC untuk memperoleh estimasi parameternya (Gamerman, 1997;



Stephens, 1997; Walsh, 2002 dan Wati, 2006). Distribusi posterior gabungan dari semua parameter model BUM-normal dinyatakan dalam persamaan (2.29).

$$p(\boldsymbol{\theta}, \mathbf{x}) = p(\mu | \sigma^2) p(\sigma^2 | \mu) p(\mathbf{x} | \sigma^2, \mu) \quad (2.29)$$

dimana:

- $p(\boldsymbol{\theta}, \mathbf{x})$  = distribusi posterior gabungan normal unimodal dari data  $\mathbf{x}$  dengan parameter  $\boldsymbol{\theta}$  yang berupa parameter  $\mu$  dan  $\sigma^2$ ,
- $p(\mu | \sigma^2)$  = distribusi prior dari parameter  $\mu$  dengan syarat parameter  $\sigma^2$  bernilai tertentu,
- $p(\sigma^2 | \mu)$  = distribusi prior dari parameter  $\sigma^2$  dengan syarat parameter  $\mu$  bernilai tertentu, dan
- $p(\mathbf{x} | \sigma^2, \mu)$  = fungsi *likelihood* dari data  $\mathbf{x}$  dengan syarat parameter  $\mu$  dan  $\sigma^2$  bernilai tertentu.

Proses selanjutnya melakukan estimasi setiap parameter pada persamaan (2.29) dengan membentuk distribusi bersyarat penuh dari setiap parameter model (Richardson dan Green, 1997 dan Dinar, 2009). Distribusi bersyarat penuh untuk masing-masing parameter dalam model BUM-normal yang didasarkan pada konsep pemodelan BMM-normal dijabarkan sebagai berikut: (Gilks, 1995 dan Astuti, 2006).

#### a. Estimasi Parameter Lokasi ( $\mu$ )

Distribusi prior dari parameter  $\mu$  adalah  $\text{normal}(\xi, K^{-1})$  dan dengan teorema Bayes berikut:

$$p(\mu | \dots) \propto \frac{p(\mu, \sigma^{-2} | \mathbf{x})}{p(\sigma^{-2} | \mathbf{x})} = \frac{p(\mathbf{x}, \mu, \sigma^{-2})}{p(\mathbf{x}, \sigma^{-2})} \propto p(\mathbf{x}, \mu, \sigma^{-2}).$$

diperoleh distribusi bersyarat penuh dari parameter lokasi  $\mu$  sebagai berikut:

$$p(\mu | \sigma^2) \propto l(\mathbf{x} | \sigma^2) p(\mu).$$

Penjabaran secara matematis dari persamaan distribusi bersyarat penuh untuk parameter lokasi sebagaimana ditunjukkan dalam persamaan (2.30) (Gils, 1995 dan Astuti, 2006).

$$\begin{aligned}
p(\mu | \dots) &\propto \exp \left\{ -\frac{\sigma^{-2}}{2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \times \exp \left\{ -\frac{K}{2} (\mu - \xi)^2 \right\} \\
&= \exp \left\{ -\frac{1}{2} \left( \left( \sigma^{-2} \sum_{i=1}^n (x_i - \mu)^2 \right) + K (\mu - \xi)^2 \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left( \sigma^{-2} \sum_{i=1}^n x_i^2 - 2\sigma^{-2} \mu \sum_{i=1}^n x_i + \sigma^{-2} \sum_{i=1}^n \mu^2 + K \mu^2 - 2K\xi\mu + K\xi^2 \right) \right\} \\
&= \exp \left\{ -\frac{1}{2} \left( \sigma^{-2} \sum_{i=1}^n x_i^2 - 2\mu \left( \sigma^{-2} \sum_{i=1}^n x_i + K\xi \right) + \mu^2 (\sigma^{-2}n + K) + K\xi^2 \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} (K + \sigma^{-2}n) \left( \mu^2 - \frac{2\mu \left( \sigma^{-2} \sum_{i=1}^n x_i + K\xi \right)}{K + \sigma^{-2}n} + \frac{\sigma^{-2} \sum_{i=1}^n x_i^2 + K\xi^2}{K + \sigma^{-2}n} \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} (K + \sigma^{-2}n) \left( \mu - \frac{\left( \sigma^{-2} \sum_{i=1}^n x_i + K\xi \right)}{K + \sigma^{-2}n} \right)^2 \right\}.
\end{aligned} \tag{2.30}$$

Bentuk terakhir dari persamaan (2.30) adalah bentuk distribusi normal, sehingga distribusi bersyarat penuh dari parameter lokasi  $\mu$  ditunjukkan dalam persamaan (2.30).

$$\mu | \sigma^2 \sim N \left( \frac{\left( \sigma^{-2} \sum_{i=1}^n x_i + K\xi \right)}{K + \sigma^{-2}n}, (K + \sigma^{-2}n)^{-1} \right). \tag{2.31}$$

#### b. Estimasi Parameter Varians ( $\sigma^{-2}$ )

Distribusi prior dari parameter  $\sigma^{-2}$  adalah  $\text{gamma}(g, h)$  sehingga bentuk distribusi bersyarat penuh untuk parameter  $\sigma^{-2}$  adalah:

$$p(\sigma^{-2} | \mu) \propto l(\mathbf{x} | \sigma^2) p(\sigma^{-2}).$$

Penjabaran secara matematis untuk distribusi bersyarat penuh dari parameter varians  $\sigma^{-2}$  dapat ditulis sebagaimana pada persamaan (2.32).

$$\begin{aligned}
p(\sigma^{-2} | \dots) &\propto (\sigma^{-2})^{\frac{n}{2}} \exp\left\{-\frac{\sigma^{-2}}{2} \sum_{i=1}^n (x_i - \mu)^2\right\} \frac{\beta^\alpha (\sigma^{-2})^{\alpha-1} \exp(-\beta \sigma^{-2})}{\Gamma(\alpha)} \\
&= \frac{\beta^\alpha (\sigma^{-2})^{\frac{n}{2} + \alpha - 1} \exp\left\{-\frac{\sigma^{-2}}{2} \sum_{i=1}^n (x_i - \mu)^2 - \beta \sigma^{-2}\right\}}{\Gamma(\alpha)} \\
&= \frac{\beta^\alpha (\sigma^{-2})^{\frac{n}{2} + \alpha - 1} \exp\left\{-\sigma^{-2} \left(\beta + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right)\right\}}{\Gamma(\alpha)} \tag{2.32} \\
&\propto \frac{\left(\beta + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right)^{\frac{n}{2} + \alpha}}{(\sigma^{-2})^{\frac{n}{2} + \alpha - 1} \exp\left\{-\sigma^{-2} \left(\beta + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right)\right\}} \\
&\quad \Gamma\left(\alpha + \frac{1}{2} n\right) \\
&\propto \Gamma\left(\alpha + \frac{1}{2} n, \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right).
\end{aligned}$$

Bentuk terakhir dari persamaan (2.32) merupakan distribusi bersyarat penuh untuk parameter varians,  $\sigma^{-2}$ , yang dapat ditulis dalam persamaan (2.33).

$$\sigma^{-2} | \mu \sim \Gamma\left(\alpha + \frac{1}{2} n, \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^2\right). \tag{2.33}$$

### c. Estimasi *Hyperparameter* ( $\beta$ )

Parameter  $\beta$  merupakan *hyperparameter* dari  $\sigma^{-2}$  yang memiliki distribusi prior  $\text{gamma}(g, h)$  sehingga bentuk distribusi bersyarat penuh dari *hyperparameter*  $\beta$  adalah:

$$p(\beta | \sigma^{-2}) \propto p(\sigma^{-2} | \alpha, \beta) p(\beta | g, h).$$

Penjabaran secara matematis bentuk distribusi bersyarat penuh dari *hyperparameter*  $\beta$  sebagaimana ditunjukkan dalam persamaan berikut:

$$\begin{aligned}
p(\beta | \dots) &\propto p(\alpha, \beta) p(\beta | g, h) \\
&= \frac{\beta^\alpha (\sigma^{-2})^{\alpha-1} \exp\{-\beta \sigma^{-2}\}}{\Gamma(\alpha)} \times \frac{h^g \beta^{g-1} \exp(-\beta h)}{\Gamma(g)} \\
&\propto \frac{\beta^{\alpha+g-1} \exp\{-\beta(h + \sigma^{-2})\} (\sigma^{-2})^{\alpha-1} h^g}{\Gamma(g + \alpha)} \\
&\propto \frac{\beta^{\alpha+g-1} \exp\{-\beta(h + \sigma^{-2})\} (h + \sigma^{-2})^{(g+\alpha)}}{\Gamma(g + \alpha)} \\
&= \Gamma(g + \alpha - 1, h + \sigma^{-2}).
\end{aligned}$$

Sehingga distribusi bersyarat penuh dari *hyperparameter*  $\beta$  dapat ditulis dalam persamaan (2.34).

$$\beta | \sigma^{-2} \sim \Gamma(g + \alpha - 1, h + \sigma^{-2}). \quad (2.34)$$

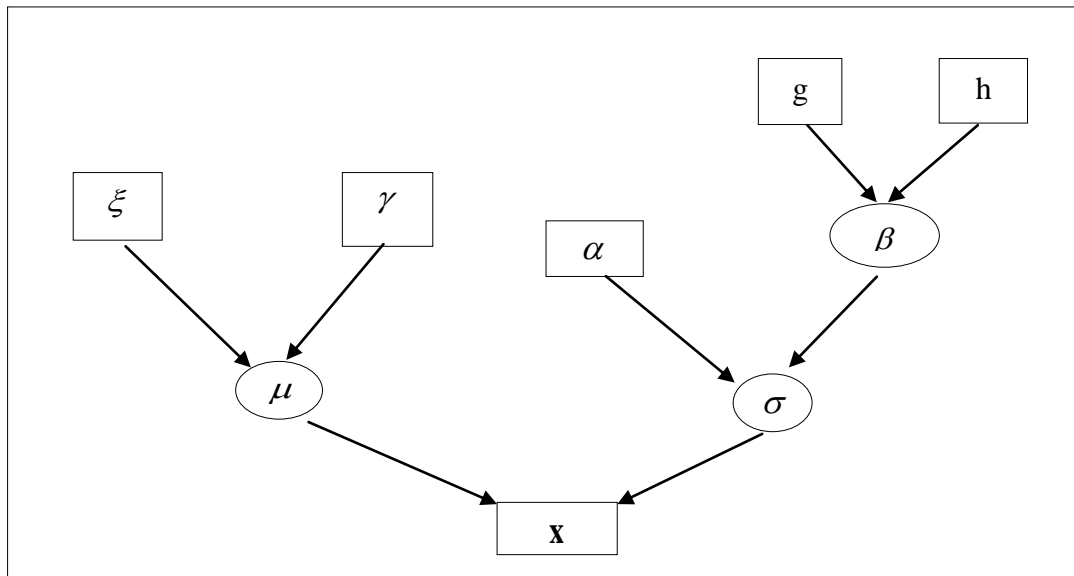
DAG untuk pemodelan BUM-normal juga disusun dalam dua *node* yang berbeda yaitu kotak dan ellips. *Node* kotak menunjukkan bahwa parameter bernilai konstan sebagai pernyataan bahwa parameter tersebut menduduki posisi hirarki yang final (tidak ada hirarki yang lebih atas lagi), sedangkan *node ellips* menunjukkan bahwa parameter bisa berubah secara *stochastic* ataupun *logical* (Iriawan, 2003b; Astuti, 2006 dan Dinar, 2009). Berikut ini akan diberikan contoh struktur hirarki model BUM-normal sebagaimana ditunjukkan dalam Gambar 2.2, dimana  $\mu$  dan  $\sigma$  menotasikan parameter komponen penyusun model BUM yang berdistribusi normal dan  $\mathbf{x}$  adalah data pengamatan yang berdistribusi *mixture* normal.

Algoritma Pemodelan BUM-normal telah terbangun dan akan dijabarkan secara lengkap pada Bab IV.

## 2.7. Model Bayesian Averaging (BMA)

Teori dan konsep dasar BMA telah banyak dibahas di dalam beberapa literatur, di antaranya Madigan dan Raftery (1994); Raftery dkk. (1997); Volinsky (1997); Hoeting dkk. (1999); Viallefont dkk. (2001); Liang dkk. (2001); Sebastiani dkk. (2006); dan Montgomery dan Nyhan (2010).

Prinsip dasar dari pemodelan BMA adalah membangun model tunggal terbaik dengan mempertimbangkan semua kemungkinan model yang bisa dibangun. Berdasarkan prinsip tersebut diharapkan dapat menghasilkan estimasi parameter model yang lebih tepat dibandingkan dengan prinsip pembentukan model tunggal yang mengabaikan ketidakpastian model. BMA merupakan solusi Bayesian yang mengatasi ketidakpastian model, dimana penyelesaian modelnya dengan merata-ratakan distribusi posterior dari semua model terbaiknya. Sebagai contoh dalam analisis model regresi linier, pemilihan prediktor yang masuk dalam model adalah tahap pemodelan yang penting. Jika ada  $p$  buah variabel prediktor, maka terdapat  $q = 2^p$  model yang bisa dibentuk dengan asumsi antar variabel prediktor tidak saling berkorelasi. Tujuan dari BMA adalah menggabungkan model-model yang tidak pasti, sehingga diperoleh satu model yang terbaik (Madigan dan Raftery, 1994; Hustienda, 2012; Purnamasari, 2011 dan Kuswanto dan Sari, 2013).



**Gambar 2.2.** Struktur Hirarki dari Prior Model BUM-Normal dalam DAG

Menurut Hoeting dkk. (1999), prediksi parameter dengan pendekatan BMA akan memanfaatkan data yang berasal dari model kombinasi hirarki. Jika  $\mathbf{M} = \{M_1, M_2, \dots, M_q\}$  adalah himpunan model yang dapat dibentuk dari  $\mathbf{M}$  dan  $\Delta$  adalah nilai yang akan diprediksi, maka prediksi BMA dimulai dengan menentukan distribusi probabilitas prior dari semua parameter model dan model

$M_s$ , dengan  $s=1,2,\dots,q$ . Model  $M_s$  diasumsikan memiliki distribusi probabilitas prior  $M_s \sim \mathcal{G}(M_s)$  dan vektor parameter model yang dihasilkan dari distribusi bersyaratnya adalah sebagai berikut:  $(\sigma^2 | M_s) \sim \mathcal{G}(\sigma^2 | M_s)$  dan untuk  $(\theta_s | \sigma^2, M_s) \sim \mathcal{G}(\theta_s | \sigma^2, M_s)$ . Berdasarkan asumsi tersebut maka data  $\mathbf{X}$  mengikuti distribusi bersyarat  $(\mathbf{x} | \theta_s, \sigma^2, M_s) \sim N(X_s \theta_s, \sigma^2 I)$ . Distribusi posterior dari  $\Delta$  jika diketahui data  $\mathbf{X}$  adalah:

$$p(\Delta | \mathbf{x}) = \sum_{s=1}^q P(\Delta | M_s, \mathbf{x}) P(M_s | \mathbf{x}), \quad (2.35)$$

dimana  $q$  merupakan banyaknya model yang mungkin terbentuk. Distribusi posterior dari  $\Delta$  jika diketahui data  $\mathbf{X}$  merupakan rata-rata dari distribusi posterior jika diketahui model diboboti oleh probabilitas model posterior. Sedangkan probabilitas posterior dari model  $M_s$  adalah:

$$p(M_s | \mathbf{x}) = \frac{p(\mathbf{x} | M_s) p(M_s)}{\sum_{s=1}^q p(\mathbf{x} | M_s) p(M_s)}, \quad (2.36)$$

dimana

$$p(\mathbf{x} | M_s) = \int p(\mathbf{x} | \theta_s, M_s) p(\theta_s | M_s) d\theta_s. \quad (2.37)$$

Persamaan (2.37) merupakan marginal *likelihood* dari model  $M_s$ ,  $p(\theta_s | M_s)$  adalah probabilitas prior dari  $\theta_s$  jika diketahui model  $M_s$ ,  $p(\mathbf{x} | \theta_s, M_s)$  adalah fungsi *likelihood* dari data  $\mathbf{X}$  dan  $p(M_s)$  merupakan probabilitas prior, jika model  $M_s$  benar. Semua probabilitas secara implisit bergantung pada model  $\mathbf{M}$ , sehingga nilai harapan dari koefisien  $\Delta$  didapatkan dengan merata-ratakan model  $\mathbf{M}$ , yaitu:

$$E(\Delta | \mathbf{x}) = \sum_{s=1}^q \hat{\Delta}_s p(M_s | \mathbf{x}),$$

dengan  $\hat{\Delta}_s = E(\Delta | \mathbf{x}, M_s)$  sehingga diperoleh persamaan untuk  $E(\Delta | \mathbf{x})$  sebagaimana ditunjukkan dalam persamaan (2.38).

$$E(\Delta | \mathbf{x}) = \sum_{s=1}^q E(\Delta | \mathbf{x}, M_s) P(M_s | \mathbf{x}), \quad (2.38)$$

dimana arti dari simbol  $P$  sama dengan  $p$ , sehingga  $P = p$

$E(\Delta | \mathbf{x})$  = nilai harapan dari koefisien  $\Delta$  bersyarat data  $\mathbf{x}$ ,  
 $P(M_s | \mathbf{x})$  = nilai pembobot pada model  $\mathbf{M}$  ke- $s$  dengan  $s = 1, 2, \dots, q$ ,  
 $E(\Delta | \mathbf{x}, M_s)$  = nilai harapan dari koefisien  $\Delta$  bersyarat data  $\mathbf{x}$  pada model  $\mathbf{M}$  ke- $s$ , dan  
 $s$  = banyaknya model terbaik yang terseleksi

Nilai  $E(\Delta | \mathbf{x})$  pada persamaan (2.38) menunjukkan nilai harapan terboboti dari  $\Delta$  di setiap model kombinasi yang mungkin (bobot ditentukan oleh prior dan model). Selanjutnya berdasarkan persamaan (2.38), ditentukan nilai varians dari  $E(\Delta | \mathbf{x})$  yang dijabarkan pada persamaan (2.39).

$$\begin{aligned}
 Var(\Delta | \mathbf{x}) &= \sum_{s=1}^q \left( \hat{\Delta}_s - E(\Delta | \mathbf{x}) \right)^2 P(M_s | \mathbf{x}) \\
 &= \sum_{s=1}^q \left( \hat{\Delta}_s^2 - 2\hat{\Delta}_s E(\Delta | \mathbf{x}) + E(\Delta | \mathbf{x})^2 \right) P(M_s | \mathbf{x}) \\
 &= \sum_{s=1}^q \left( \text{var}[\Delta | \mathbf{x}, M_s] + \hat{\Delta}_s^2 \right) P(M_s | \mathbf{x}) - E(\Delta | \mathbf{x})^2 \\
 &= \sum_{s=1}^q \left( \text{var}[\Delta | \mathbf{x}, M_s] + E[\Delta | \mathbf{x}, M_s]^2 \right) P(M_s | \mathbf{x}) - E(\Delta | \mathbf{x})^2,
 \end{aligned} \quad (2.39)$$

dengan  $\hat{\Delta}_s = E(\Delta | \mathbf{x}, M_s)$ ,

dimana:

$Var(\Delta | \mathbf{x})$  = nilai varians dari koefisien  $\Delta$  bersyarat data  $\mathbf{x}$ ,  
 $\text{var}[\Delta | \mathbf{x}, M_s]$  = nilai varians dari koefisien  $\Delta$  bersyarat data  $\mathbf{x}$  dan model  $\mathbf{M}$  ke- $s$  dengan  $s = 1, 2, \dots, q$ ,  
 $[E(\Delta | \mathbf{x}, M_s)]^2$  = nilai harapan kuadrat dari koefisien  $\Delta$  bersyarat data  $\mathbf{x}$  pada model  $\mathbf{M}$  ke- $s$ ,  
 $P(M_s | \mathbf{x})$  = nilai pembobot pada model  $\mathbf{M}$  ke- $s$ ,  
 $E[\Delta | \mathbf{x}]^2$  = nilai harapan kuadrat dari koefisien  $\Delta$  bersyarat data  $\mathbf{x}$ , dan  
 $s$  = banyaknya model terbaik yang terseleksi

### 2.7.1. Model Bayesian *Averaging* Normal Unimodal (BMA-Normal Unimodal)

Pemodelan BMA-normal unimodal dilakukan untuk data yang memiliki distribusi normal unimodal sebagaimana konsep yang digunakan dalam pemodelan BUM-normal. Pemodelan BMA-normal unimodal dibangun berdasarkan sejumlah model BUM-normal yang telah terbentuk. Dalam hal ini, model BMA-normal unimodal dibangun dari sebanyak 1.000 model BUM-normal yang telah memenuhi asumsi pemodelan Bayesian. Konsep pembentukan 1.000 model BUM-normal sebagaimana telah dijabarkan sebelumnya dengan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler*. Untuk distribusi bersyarat penuh dari parameter-parameter model normal unimodal juga telah dijabarkan pada sub bab 2.5.

Untuk mengimplementasikan persamaan (2.38) (Hoeting dkk., 1999) pada kasus data berdistribusi normal unimodal dari sebanyak 1.000 model, maka model BMA-normal unimodal akan dibentuk berdasarkan persamaan berikut:

$$E(\Delta | \mathbf{x}) = \sum_{s=1}^q P(M_s | \mathbf{x}) E(\Delta | M_s, \mathbf{x}),$$

dimana  $E(\Delta | \mathbf{x})$  merupakan model BMA-normal unimodal yang diharapkan dari data  $\mathbf{x}$  dan parameter model  $\Delta$ .  $P(M_s | \mathbf{x})$  adalah nilai pembobot pada model ke- $s$  dalam  $\mathbf{M}$  dengan  $s = 1, 2, \dots, q$ , dan  $s$  merupakan jumlah model terbaik yang terseleksi di antara  $q$  model yang bisa dibangun dalam pemodelan dan masing-masing model memiliki distribusi normal unimodal dari data  $\mathbf{x}$  dan  $E(\Delta | M_s, \mathbf{x})$  adalah model prediksi ke- $s$  dalam  $\mathbf{M}$  yang berdistribusi normal unimodal dari data  $\mathbf{x}$  dengan parameter model  $\Delta$ .

Untuk memperoleh model prediksi  $M_s$  pada  $E(\Delta | M_s, \mathbf{x})$  memanfaatkan Algoritma MCMC pendekatan *Gibbs Sampler* dengan jumlah iterasi parameter yang dibangkitkan sebanyak 5.000 kali dan jumlah model yang dibangkitkan sebanyak 1.000. Selanjutnya untuk memperoleh nilai pembobot pada masing-masing model  $\mathbf{M}$  ke- $s$  digunakan persamaan (2.40) (Hustianda, 2012).



$$P(M_s | \mathbf{x}) = \frac{e^{-0,5*(BIC_s - \text{maksimum} BIC)}}{\sum_{s=1}^q e^{-0,5*(BIC_s - \text{maksimum} BIC)}}, \text{ dengan } s = 1, 2, \dots, q, \quad (2.40)$$

dimana rumus BIC (*Bayesian Information Criterion*) dapat dijabarkan dalam persamaan (2.41).

$$BIC = -2 \ln L + v \ln(n) = n \ln(\hat{\sigma}_\varepsilon^2) + v \ln(n), \quad (2.41)$$

dengan  $\hat{\sigma}_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})^2$ ,

dimana

- $L$  = fungsi *likelihood* normal unimodal sebagaimana pada persamaan (2.28) ,
- $n$  = ukuran sampel data  $\mathbf{X}$  ,
- $v$  = jumlah parameter model, dan
- $\hat{\sigma}_\varepsilon^2$  = varians dari *error* model BUM-normal.

Tujuan utama dari pemodelan BMA-normal unimodal adalah menentukan model tunggal terbaik berdasarkan sejumlah model BUM-normal yang dibangkitkan. Untuk memperoleh model tunggal BMA-normal unimodal terbaik tersebut dibutuhkan sebuah metode seleksi model yang salah satunya adalah metode seleksi *Occam's Window* (Madigan dan Raftery, 1994; Hoeting dkk., 1999 dan Hustienda, 2012).

### 2.7.2. Seleksi Model dengan Metode Occam's Window

Berdasarkan konsep BMA, akan dipilih model yang akan masuk dalam persamaan (2.38) dari sejumlah  $q$  model yang mungkin terbentuk. Ada beberapa metode untuk menyeleksi model yang masuk dalam persamaan (2.38) berdasarkan probabilitas posteriornya, di antaranya adalah metode *Occam's Window* (OW) dari Madigan dan Raftery (1994). Metode *Occam's Window* ini banyak digunakan dalam penelitian yang terkait dengan pemodelan BMA dan hasil yang diperoleh cukup baik dalam memilih model BMA terbaik (Madigan dan Raftery, 1994 dan Hustienda, 2012).

Metode *Occam's Window* digunakan untuk menyeleksi model yang masuk dalam rumus BMA berdasarkan probabilitas posteriornya. Model yang diterima

dengan metode ini (model yang bisa masuk dalam pemodelan BMA) harus memenuhi persamaan (2.42) (Madigan dan Raftery, 1994 dan Hustienda, 2012).

$$\mathcal{A}' = \{M_s : \frac{\text{maksimum}_l(P(M_l | \mathbf{x}))}{P(M_s | \mathbf{x})} \leq c\} \quad (2.42)$$

dimana  $\mathcal{A}'$  merupakan posterior odds model ke- $s$ . Menurut Jeffreys (1961), batasan nilai  $c$  sebesar 20 merupakan nilai batas seleksi model ke- $s$  yang bisa masuk atau tidak dalam pemodelan BMA. Hal ini setara dengan daerah penerimaan dan penolakan hipotesis dengan taraf nyata  $\alpha = 5\%$  jika menggunakan kriteria uji melalui  $p$ -value (Jeffreys, 1961). Kass dan Raftery (1995) dalam Wati (2006) juga menjelaskan tentang nilai 20 dalam konsep *Bayes Factor* (BF) melalui tabulasi nilai BF untuk memudahkan interpretasi. Nilai BF sebesar 20 menunjukkan bahwa model distribusi data yang dihipotesiskan dinyatakan positif sesuai untuk data yang diamati.

Jika suatu model memiliki nilai  $\mathcal{A}'$  lebih besar dari  $c=20$ , maka model tersebut tidak masuk dalam perhitungan BMA dan harus dikeluarkan dari persamaan (2.38) dan sebaliknya jika suatu model memiliki nilai  $\mathcal{A}'$  lebih kecil atau sama dengan  $c=20$ , maka model tersebut masuk dalam perhitungan BMA dan harus diikutsertakan dalam persamaan (2.38). Rumus  $\text{maksimum}_l(P(M_l | \mathbf{x}))$  pada persamaan (2.42) merupakan model ke- $l$  dalam  $\mathbf{M}$  yang memiliki nilai probabilitas posterior tertinggi. Probabilitas posterior dari setiap parameter model yang signifikan ditentukan dengan merata-ratakan probabilitas posterior dari setiap parameter dari sebanyak  $s$  model terbaik yang terpilih (Madigan dan Raftery, 1994 dan Hustienda, 2012). Berdasarkan hasil seleksi model dengan metode Occam's Window ini, maka akan diperoleh model BMA-normal unimodal terbaik melalui persamaan (2.43).

$$E(\Delta | \mathbf{x}) = \sum_{s=1}^q P(M_s | \mathbf{x}) E(\Delta | M_s, \mathbf{x}), \quad (2.43)$$

dengan nilai varians dihitung dengan rumus persamaan (2.44).

$$\text{Var}(\Delta | \mathbf{x}) = \sum_{s=1}^q (\text{var}(\Delta | \mathbf{x}, M_s) + [E(\Delta | M_s, \mathbf{x})^2] P(M_s | \mathbf{x}) - E(\Delta | \mathbf{x})^2). \quad (2.44)$$

Penjabaran dari persamaan (2.43) dan persamaan (2.44) sebagaimana telah ditunjukkan pada persamaan (2.38) dan persamaan (2.39). Algoritma Pemodelan

BMA-normal unimodal telah terbangun dan secara lengkap dijabarkan secara lengkap pada Bab V.

## 2.8. Uji Kolmogorov-Smirnov (KS)

Uji Kolmogorov-Smirnov (KS) merupakan suatu uji yang digunakan untuk mengetahui apakah suatu data kontinyu mengikuti distribusi tertentu atau apakah suatu distribusi dugaan benar-benar telah sesuai dengan data yang diamati (Iriawan, 2003a). Menurut Iriawan (2003a), konsep teori dari uji KS adalah membandingkan fungsi distribusi kumulatif (CDF) empiris yaitu  $F_n(x_i)$  dengan fungsi distribusi kumulatif hipotesis (CDF dugaan), yaitu  $\hat{F}(x)$ . Jika  $X_1, X_2, \dots, X_n$  adalah order statistik dari variabel random yang independen dengan distribusi kumulatif hipotesis  $\hat{F}(x)$  dan distribusi kumulatif empiris (observasi) didefinisikan dalam persamaan (2.45)

$$F_n(x_i) = \frac{\text{Banyaknya Data } X_i \leq x_i}{n} \text{ untuk } i=1, 2, 3, \dots, m < n \quad (2.45)$$

dimana  $F_n(x_i)$  adalah fungsi step kontinyu kanan (*right continuous step function*). Setiap nilai  $F_n(x_i)$  akan dibandingkan dengan nilai distribusi hipotesis (dugaan)  $\hat{F}(x)$ , sehingga statistik uji yang digunakan dalam uji Kolmogorov-Smirnov adalah jarak vertikal terbesar (maksimum) antara  $F_n(x_i)$  dan  $\hat{F}(x)$  yang dinotasikan dengan  $D_n$  (Iriawan, 2003a dan Sheskin, 2007). Menurut Iriawan (2003a) dan Dinar (2009), uji hipotesis yang digunakan dalam konsep teori Kolmogorov-Smirnov adalah sebagai berikut:

$H_0$ : Data  $X$  adalah variabel random independen yang berdistribusi sesuai dengan distribusi  $\hat{F}(x)$

$H_1$ : Data  $X$  adalah variabel random independen yang berdistribusi tidak sesuai dengan distribusi  $\hat{F}(x)$ .

Penolakan hipotesis dilakukan menggunakan statistik uji  $D_n$  yang dapat dituliskan dalam persamaan (2.46).

$$D_n = \sup |F_n(x) - \hat{F}(x)| \quad (2.46)$$

Berdasarkan nilai statistik uji  $D_n$  pada persamaan (2.46) dapat diketahui apakah hipotesis nol ditolak atau tidak. Jika nilai  $D_n > d_n$  atau nilai  $P\text{-value} < 0,05$ , maka hipotesis nol ( $H_0$ ) ditolak dan sebaliknya jika  $D_n \leq d_n$  atau nilai  $P\text{-value} \geq 0,05$ , maka hipotesis nol ( $H_0$ ) gagal ditolak. Nilai  $d_n$  merupakan nilai acuan yang diperoleh dari tabel Kolmogorov-Smirnov.

Distribusi kumulatif hipotesis  $\hat{F}(x)$  pada persamaan (2.46) akan disesuaikan dengan pemodelan yang sedang dilakukan. Apabila data dimodelkan dengan BUM-normal, maka distribusi kumulatif  $\hat{F}(x)$  yang digunakan adalah distribusi kumulatif untuk model BUM-normal. Dalam hal ini distribusi kumulatif  $\hat{F}(x)$  dapat berupa model BUM-normal, model BMA-normal unimodal, model BMM-normal, dan model BMMA-normal, sesuai dengan tiap bahasan di bab-bab selanjutnya. Menurut Iriawan (2003a), pemilihan uji Kolmogorov-Smirnov dilakukan karena:

1. Distribusi yang akan diuji adalah distribusi kontinyu
2. Tes dari uji ini berlaku untuk sembarang ukuran sampel ( $n$ )
3. Tes dari uji ini telah dilakukan modifikasi sehingga untuk distribusi eksponensial, normal dan Weibull diperbolehkan dengan menggunakan dugaan nilai parameter-parameternya (Law dan Kelton, 2000).

Konsep teori KS ini juga dapat digunakan untuk menguji apakah dua sampel independen berasal dari satu populasi yang sama atau berasal dari populasi yang berbeda. Hal ini dilakukan dengan memperhatikan kesesuaian antara dua distribusi kumulatif dari kedua himpunan sampel independen yang diamati (Sheskin, 2007).

Ilustrasi konsep teori uji Kolmogorov-Smirnov seperti disajikan dalam Gambar 2.3.

## **2.9. Highest Probability Distribution (HPD) dengan Credible Interval ( $1-\alpha$ )**

HPD merupakan sebuah konsep densitas setimbang untuk data yang berpola tidak simetri (asimetri), dimana cara pembentukannya didasarkan atas konsep densitas simetri yang merupakan cara paling mudah dalam mempelajari

pembentukan interval kepercayaan (*credible interval*) dari data (Iriawan, 2012). Pada kondisi pola data tidak simetri, lokasi rata-rata, median dan modus terletak pada posisi yang berbeda-beda, sehingga pembentukan interval kepercayaan untuk densitas pola simetri tidak dapat digunakan.

Untuk membentuk *credible interval* pola data tidak simetri (bisa miring ke kanan atau ke kiri) dengan HPD digunakan acuan lokasi modus. Jarak lompatan menuju batas atas dan batas bawah *credible interval* tidak sama, tetapi ketinggian densitas di dua batas tersebut dan luasan di bawah kurva densitas di antara nilai-nilai batas tersebut harus tetap terjamin sebesar  $(1-\alpha)$ . Menurut Iriawan (2012), untuk kemudahan pemahaman konsep densitas setimbang dapat dijelaskan melalui sebuah algoritma yang disebut Algoritma Pembentukan Densitas Setimbang HPD. Tahapan-tahapan algoritma ini secara lengkap disajikan dalam Algoritma 2.3.

**Algoritma 2.3.** Algoritma Pembentukan Densitas Setimbang HPD

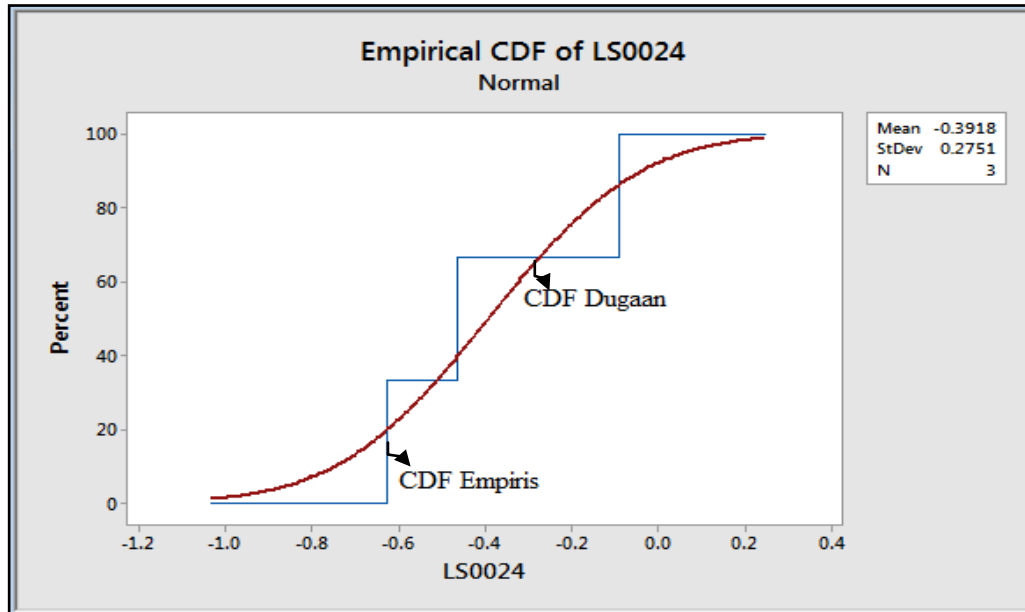
Apabila diasumsikan telah diketahui pola data pengamatan yang berpola tidak simetris dengan densitas  $f(x|\theta)$ , maka batas-batas *credible interval* dari  $x$  dapat ditentukan dengan tahapan proses sebagai berikut:

Langkah 1. Tentukan modus dari densitas  $f(x|\theta)$ ,

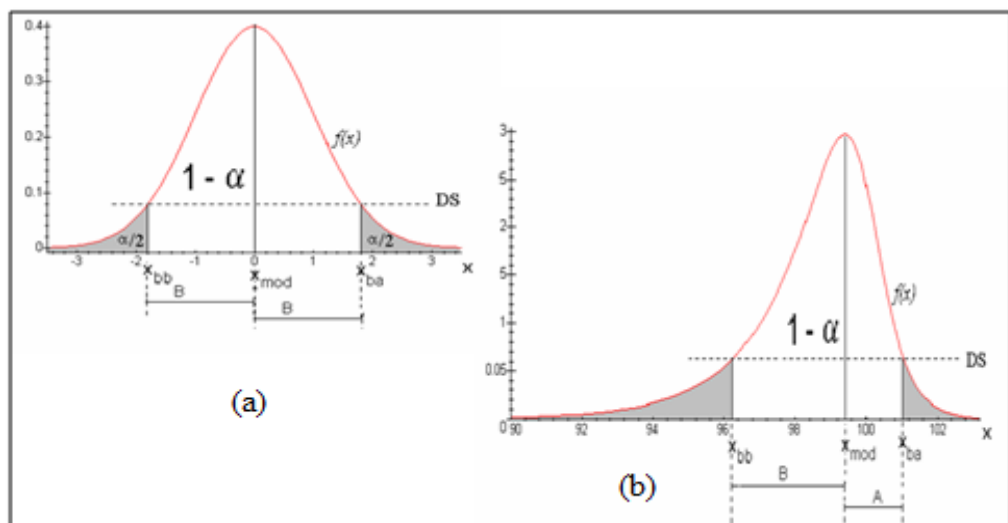
Langkah 2. Letakkan pusat data pada modus hasil Langkah 1.

Langkah 3. Buat lompatan langkah ke kiri dan ke kanan dari modus untuk menentukan nilai batas bawah (namakan  $a$ ) dan nilai batas atas (namakan  $b$ ) dengan tetap mengontrol kesetimbangan nilai densitasnya, yaitu  $f(a) = f(b)$  untuk memperoleh *credible interval* sebesar  $(1-\alpha)$ , yang dapat ditulis sebagai  $P(a < x < b) = \int_a^b f(x)dx = (1-\alpha)$  dengan  $x$  adalah data pengamatan individu.

Konsep densitas setimbang untuk pola data simetri dan tidak simetri diberikan dalam Gambar 2.4 (Iriawan, 2001).



**Gambar 2.3.** Ilustrasi Konsep Teori Uji Kolmogorov-Smirnov (KS)



**Gambar 2.4.** Konsep Densitas Setimbang untuk Pola Data Simetri dan Tidak Simetri, (a) untuk *Confidence Interval* (b) untuk *Credible Interval* (Iriawan, 2001)

Catatan: DS adalah densitas setimbang

$x_{mod}$  adalah modulus

$x_{ba} = b$  adalah batas atas interval

$x_{bb} = a$  adalah batas bawah interval

A adalah jarak dari  $x_{mod}$  ke  $x_{ba}$

B adalah jarak dari  $x_{mod}$  ke  $x_{bb}$

*Halaman ini sengaja dikosongkan*

## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data *Chickpea* dalam penelitian Harijati (2007) yang selanjutnya akan disebut data *Chickpea* dalam penelitian ini. Struktur data *Chickpea* yang dijabarkan dalam Tabel 3.1 merupakan data perbedaan ekspresi gen dari jaringan tanaman *Chickpea* jenis *Lasseter* (Gambar 3.1) yang ditanam di Indonesia yang merupakan hasil percobaan *microarray*. Tahapan-tahapan teknologi *microarray* yang digunakan sebagaimana ditunjukkan pada Gambar 3.2, yaitu ekstraksi RNA untuk membentuk cDNA dilakukan pada tanaman *Chickpea* pada dua kondisi sakit dan sehat. Selanjutnya melakukan pelabelan cDNA-Cy3 (label untuk jaringan tanaman kondisi sehat) dan pelabelan cDNA-Cy5 (label untuk jaringan tanaman kondisi sakit) dengan pewarna pendaran (*fluorescent*). Langkah berikutnya mengawinkan silang (*hybridization*) hasil pelabelan itu ke suatu *array* pada *probe* sehingga dapat diketahui nilai ekspresi dari masing-masing ID gen dari jaringan tanaman *Chickpea*. Hasil ekspresi gen yang diperoleh jumlahnya ribuan (kurang lebih 4.000 ID gen), tetapi ID gen yang dapat diidentifikasi ekspresinya dengan baik hanya sejumlah 62 ID gen (Harijati, 2007).

Jenis tanaman *Chickpea* yang diamati pada penelitian Harijati (2007) ada dua, yaitu *Lasseter* (jenis yang tidak tahan terhadap penyakit) dan *FLIP94508C* (jenis yang tahan terhadap penyakit). Tiga waktu pengamatan terhadap serangan penyakit yang diamati yaitu 24 jam, 48 jam dan 72 jam setelah serangan. Pengamatan terhadap masing-masing perlakuan diulang sebanyak tiga kali secara bebas (*independent*). Berdasarkan struktur data yang disajikan pada Tabel 3.1, data perbedaan ekspresi gen yang akan digunakan dalam penelitian ini hanya pada jenis tanaman *Chickpea Lasseter* muda yang diserang penyakit jamur patogen *Ascochyta Rabiei* selama 24 jam (LY-24) pada 26 ID gen yang memiliki fungsi ketahanan sebanyak 15 ID gen dan 11 ID gen dengan fungsi energi. Dalam penelitian ini data yang dianalisis hanya pada dua fungsi ID gen dari sembilan fungsi ID gen yang teramati (Lampiran 2). Hal ini dilakukan karena dua kelompok



ID gen dengan fungsi ketahanan dan fungsi energi ini merupakan kelompok ID gen yang sangat berperan penting terkait dengan serangan penyakit jamur patogen *Ascochyta Rabiei* (Harijati, 2007), selain itu dengan 26 ID gen yang diamati tersebut juga telah menjawab tantangan penelitian yang dilakukan. Data selainnya akan digunakan untuk penelitian lebih lanjut. Secara lengkap, fungsi-fungsi ID gen yang teramati dalam jaringan tanaman *Chickpea* (Harijati, 2007) dijabarkan dalam Lampiran 2.

Bentuk data yang dianalisis dan disajikan dalam penelitian ini adalah data  $\ln$  rasio dari nilai ekspresi gen kondisi sakit terhadap nilai ekspresi gen kondisi

sehat yaitu:  $\ln\left(\frac{\text{nilai ekspresi gen sakit}}{\text{nilai ekspresi gen sehat}}\right)$  yang berasal dari dua sampel

berpasangan, yaitu antara sampel ID gen pada tanaman *Chickpea* yang dikondisikan sakit dan ID gen yang dikondisikan sehat untuk setiap ID gen yang diamati dengan pengulangan sebanyak tiga kali. Kondisi berpasangan ini telah memenuhi konsep/kaidah dua sampel berpasangan karena suatu ID gen yang sama diamati pada dua keadaan yang berbeda. Sebagai contoh hasil pengamatan data

*Chickpea* dalam  $\ln\left(\frac{\text{nilai ekspresi gen sakit}}{\text{nilai ekspresi gen sehat}}\right)$  untuk ID gen LS0024 ulangan 1

(U1), ulangan 2 (U2) dan ulangan 3 (U3) pada fungsi ketahanan adalah -0.6774, -0.4635 dan -0.3918 secara berturut-turut.

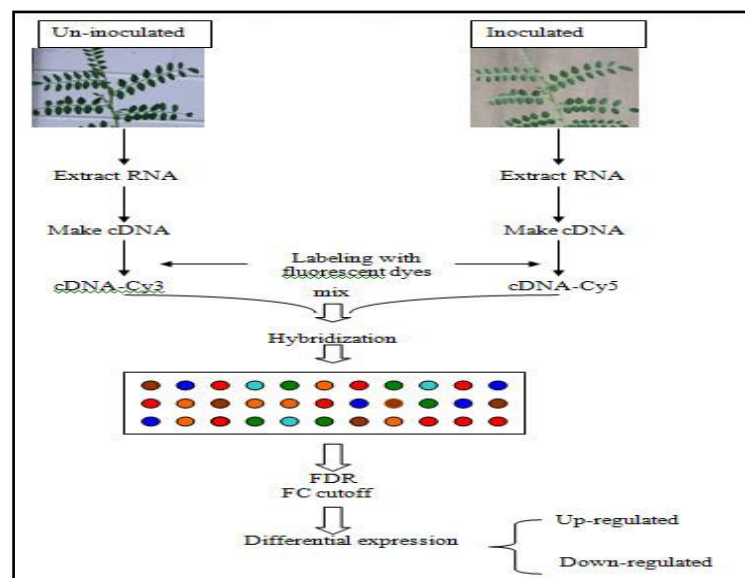
Distribusi data *Chickpea* dalam penelitian ini menggunakan pendekatan distribusi normal berdasarkan informasi hasil penelitian-penelitian sebelumnya yang terkait dengan data *microarray* (Schena dkk., 1995; Schena dkk., 1996; DeRisi dkk., 1996; Thomas dkk., 2001; Lonnstedt dan Speed, 2002; Tusher dkk., 2003; Giles dan Kipling, 2003; Harijati, 2007). Pendekatan distribusi kenormalan data dilakukan bukan berdasarkan hasil statistik uji kenormalan karena ukuran sampel yang tersedia sangat sedikit ( $n=3$ ). Deskripsi dan eksplorasi data yang digunakan dalam penelitian ini disajikan dalam Bab IV dan pernyataan penggunaan keslian data dari sumber aslinya telah dilampirkan pada Lampiran 8.

### 3.2. Program Komputasi dan *Software* Penelitian

Dalam penelitian ini dibangun algoritma-algoritma komputasi yang terkait dengan pemodelan yang akan dibangun. Program komputasi dan *software* yang digunakan untuk mengimplementasikan algoritma-algoritma yang telah dibangun dalam penelitian ini adalah *Software R* yang *compatible* dengan *Software WinBUGS* (R-WinBUGS). Penggunaan *Software R* karena *software* ini bersifat *open source* dan memiliki fasilitas pemrograman yang sangat komplit.



**Gambar 3.1.** Tanaman *Chickpea* Jenis *Lasseter* pada Umur Muda dan Umur Tua



**Gambar 3.2.** Tahapan Identifikasi Ekspresi Gen Jaringan Tanaman *Chickpea* dengan Teknologi *Microarray*

**Tabel 3.1.** Struktur Data Lengkap Pengukuran Perbedaan Ekspresi Gen Hasil Penelitian Harijati (2007)

No.	ID Gen	Lasseter (Tidak Tahan Penyakit)						FLIP508 (Tahan Penyakit)					
		Muda			Tua			Muda			Tua		
		24J	48J	72J	24J	48J	72J	24J	48J	72J	24J	48J	72J
		Ulangan	Ulangan	Ulangan	Ulangan	Ulangan	Ulangan	Ulangan	Ulangan	Ulangan	Ulangan	Ulangan	Ulangan
		U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3	U1 U2 U3
1	LS0024	REG <sub>(1)1111</sub> REG <sub>(1)1112</sub> REG <sub>(1)1113</sub>	REG <sub>(1)1121</sub> REG <sub>(1)1122</sub> REG <sub>(1)1123</sub>	REG <sub>(1)1131</sub> REG <sub>(1)1132</sub> REG <sub>(1)1133</sub>	REG <sub>(1)1211</sub> REG <sub>(1)1212</sub> REG <sub>(1)1213</sub>	REG <sub>(1)1221</sub> REG <sub>(1)1222</sub> REG <sub>(1)1223</sub>	REG <sub>(1)1231</sub> REG <sub>(1)1232</sub> REG <sub>(1)1233</sub>	REG <sub>(1)2111</sub> REG <sub>(1)2112</sub> REG <sub>(1)2113</sub>	REG <sub>(1)2121</sub> REG <sub>(1)2122</sub> REG <sub>(1)2123</sub>	REG <sub>(1)2131</sub> REG <sub>(1)2132</sub> REG <sub>(1)2133</sub>	REG <sub>(1)2211</sub> REG <sub>(1)2212</sub> REG <sub>(1)2213</sub>	REG <sub>(1)2221</sub> REG <sub>(1)2222</sub> REG <sub>(1)2223</sub>	REG <sub>(1)2231</sub> REG <sub>(1)2232</sub> REG <sub>(1)2233</sub>
2	LS0035	REG <sub>(2)1111</sub> REG <sub>(2)1112</sub> REG <sub>(2)1113</sub>	REG <sub>(2)1121</sub> REG <sub>(2)1122</sub> REG <sub>(2)1123</sub>	REG <sub>(2)1131</sub> REG <sub>(2)1132</sub> REG <sub>(2)1133</sub>	REG <sub>(2)1211</sub> REG <sub>(2)1212</sub> REG <sub>(2)1213</sub>	REG <sub>(2)1221</sub> REG <sub>(2)1222</sub> REG <sub>(2)1223</sub>	REG <sub>(2)1231</sub> REG <sub>(2)1232</sub> REG <sub>(2)1233</sub>	REG <sub>(2)2111</sub> REG <sub>(2)2112</sub> REG <sub>(2)2113</sub>	REG <sub>(2)2121</sub> REG <sub>(2)2122</sub> REG <sub>(2)2123</sub>	REG <sub>(2)2131</sub> REG <sub>(2)2132</sub> REG <sub>(2)2133</sub>	REG <sub>(2)2211</sub> REG <sub>(2)2212</sub> REG <sub>(2)2213</sub>	REG <sub>(2)2221</sub> REG <sub>(2)2222</sub> REG <sub>(2)2223</sub>	REG <sub>(2)2231</sub> REG <sub>(2)2232</sub> REG <sub>(2)2233</sub>
3	LS0616	REG <sub>(3)1111</sub> REG <sub>(3)1112</sub> REG <sub>(3)1113</sub>	REG <sub>(3)1121</sub> REG <sub>(3)1122</sub> REG <sub>(3)1123</sub>	REG <sub>(3)1131</sub> REG <sub>(3)1132</sub> REG <sub>(3)1133</sub>	REG <sub>(3)1211</sub> REG <sub>(3)1212</sub> REG <sub>(3)1213</sub>	REG <sub>(3)1221</sub> REG <sub>(3)1222</sub> REG <sub>(3)1223</sub>	REG <sub>(3)1231</sub> REG <sub>(3)1232</sub> REG <sub>(3)1233</sub>	REG <sub>(3)2111</sub> REG <sub>(3)2112</sub> REG <sub>(3)2113</sub>	REG <sub>(3)2121</sub> REG <sub>(3)2122</sub> REG <sub>(3)2123</sub>	REG <sub>(3)2131</sub> REG <sub>(3)2132</sub> REG <sub>(3)2133</sub>	REG <sub>(3)2211</sub> REG <sub>(3)2212</sub> REG <sub>(3)2213</sub>	REG <sub>(3)2221</sub> REG <sub>(3)2222</sub> REG <sub>(3)2223</sub>	REG <sub>(3)2231</sub> REG <sub>(3)2232</sub> REG <sub>(3)2233</sub>
4	LS0159	REG <sub>(4)1111</sub> REG <sub>(4)1112</sub> REG <sub>(4)1113</sub>	REG <sub>(4)1121</sub> REG <sub>(4)1122</sub> REG <sub>(4)1123</sub>	REG <sub>(4)1131</sub> REG <sub>(4)1132</sub> REG <sub>(4)1133</sub>	REG <sub>(4)1211</sub> REG <sub>(4)1212</sub> REG <sub>(4)1213</sub>	REG <sub>(4)1221</sub> REG <sub>(4)1222</sub> REG <sub>(4)1223</sub>	REG <sub>(4)1231</sub> REG <sub>(4)1232</sub> REG <sub>(4)1233</sub>	REG <sub>(4)2111</sub> REG <sub>(4)2112</sub> REG <sub>(4)2113</sub>	REG <sub>(4)2121</sub> REG <sub>(4)2122</sub> REG <sub>(4)2123</sub>	REG <sub>(4)2131</sub> REG <sub>(4)2132</sub> REG <sub>(4)2133</sub>	REG <sub>(4)2211</sub> REG <sub>(4)2212</sub> REG <sub>(4)2213</sub>	REG <sub>(4)2221</sub> REG <sub>(4)2222</sub> REG <sub>(4)2223</sub>	REG <sub>(4)2231</sub> REG <sub>(4)2232</sub> REG <sub>(4)2233</sub>
5	LS0162	REG <sub>(5)1111</sub> REG <sub>(5)1112</sub> REG <sub>(5)1113</sub>	REG <sub>(5)1121</sub> REG <sub>(5)1122</sub> REG <sub>(5)1123</sub>	REG <sub>(5)1131</sub> REG <sub>(5)1132</sub> REG <sub>(5)1133</sub>	REG <sub>(5)1211</sub> REG <sub>(5)1212</sub> REG <sub>(5)1213</sub>	REG <sub>(5)1221</sub> REG <sub>(5)1222</sub> REG <sub>(5)1223</sub>	REG <sub>(5)1231</sub> REG <sub>(5)1232</sub> REG <sub>(5)1233</sub>	REG <sub>(5)2111</sub> REG <sub>(5)2112</sub> REG <sub>(5)2113</sub>	REG <sub>(5)2121</sub> REG <sub>(5)2122</sub> REG <sub>(5)2123</sub>	REG <sub>(5)2131</sub> REG <sub>(5)2132</sub> REG <sub>(5)2133</sub>	REG <sub>(5)2211</sub> REG <sub>(5)2212</sub> REG <sub>(5)2213</sub>	REG <sub>(5)2221</sub> REG <sub>(5)2222</sub> REG <sub>(5)2223</sub>	REG <sub>(5)2231</sub> REG <sub>(5)2232</sub> REG <sub>(5)2233</sub>
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.
62	U072	REG <sub>(4704)1111</sub> REG <sub>(4704)1112</sub> REG <sub>(4704)1113</sub>	REG <sub>(4704)1121</sub> REG <sub>(4704)1122</sub> REG <sub>(4704)1123</sub>	REG <sub>(4704)1131</sub> REG <sub>(4704)1132</sub> REG <sub>(4704)1133</sub>	REG <sub>(4704)1211</sub> REG <sub>(4704)1212</sub> REG <sub>(4704)1213</sub>	REG <sub>(4704)1221</sub> REG <sub>(4704)1222</sub> REG <sub>(4704)1223</sub>	REG <sub>(4704)1231</sub> REG <sub>(4704)1232</sub> REG <sub>(4704)1233</sub>	REG <sub>(4704)2111</sub> REG <sub>(4704)2112</sub> REG <sub>(4704)2113</sub>	REG <sub>(4704)2121</sub> REG <sub>(4704)2122</sub> REG <sub>(4704)2123</sub>	REG <sub>(4704)2131</sub> REG <sub>(4704)2132</sub> REG <sub>(4704)2133</sub>	REG <sub>(4704)2211</sub> REG <sub>(4704)2212</sub> REG <sub>(4704)2213</sub>	REG <sub>(4704)2221</sub> REG <sub>(4704)2222</sub> REG <sub>(4704)2223</sub>	REG <sub>(4704)2231</sub> REG <sub>(4704)2232</sub> REG <sub>(4704)2233</sub>

Keterangan: U=Ulangan, J=Jam, REG=Rasio Ekspresi Gen (LogBase2(nilai EG sakit/nilai EG sehat)), EG=Ekspresi Gen. Data yang dianalisis dan yang digunakan dalam penelitian ini hanya data *Lasseter* Muda 24J Ulangan U1, U2, dan U3 untuk 26 ID Gen.

### 3.3. Metode Penelitian

Berdasarkan tujuan penelitian yang ingin dicapai dalam penelitian ini, yaitu membentuk model BMMA untuk kasus data simulasi dan data *Chickpea* tentang perbedaan ekspresi gen, sehingga dalam metode penelitian ini disajikan berbagai tahapan pembangkitan data simulasi dan pemodelan. Secara berturut-turut, pemodelan yang dilakukan adalah Bayesian Unimodal Model (BUM), Bayesian Model *Averaging* (BMA), Bayesian *Mixture* Model (BMM), dan terakhir Bayesian *Mixture* Model *Averaging* (BMMA). Tahapan-tahapan pembangkitan data simulasi dan pemodelan data dijabarkan sebagai berikut:

#### a. Pembangkitan Data Simulasi

Pembangkitan data simulasi didasarkan pada data *Chickpea* dengan konsep data *driven* dengan ukuran sampel  $n = 3$ . Dalam simulasi ini, penggunaan distribusi normal unimodal untuk data *Chickpea* dilakukan sesuai dengan informasi awal yang telah diperoleh berdasarkan penelitian-penelitian sebelumnya dan untuk kemudahan pendekatan distribusi data. Oleh karena itu, pembangkitan data simulasi dilakukan untuk melihat sejauh mana variabilitas data yang bisa dibangkitkan dengan tetap mengikuti pola data *Chickpea*. Selain itu juga ingin membuktikan apakah distribusi data *Chickpea* masih tetap mengikuti distribusi normal unimodal saja (sesuai informasi awal tentang data *Chickpea* yang diketahui) ataukah mengikuti distribusi *mixture*. Pembangkitan data simulasi ini di skenario tertentu dan disajikan dalam suatu algoritma agar diperoleh data simulasi yang sepola dengan data *Chickpea* dimana hal ini belum pernah dilakukan Harijati (2007). Tahapan-tahapan pembangkitan data simulasi adalah sebagai berikut:

1. Pandang data  $\mathbf{x}$  sebagai variabel acak dengan  $x$  sebagai data pengamatan univariat dari data *Chickpea* dengan ukuran sampel  $n = 3$ , sehingga data pengamatannya adalah  $x_1, x_2, x_3$ .
2. Urutkan data pengamatan  $x$  sehingga diperoleh  $x_{(1)}, x_{(2)}, x_{(3)}$ .
3. Hitung rata-rata ( $\bar{x}$ ) dan simpangan baku ( $s_x$ ) dari data  $\mathbf{x}$ .

4. Tetapkan toleransi selisih ( $\varepsilon$ ) sebesar 0,0999 (penetapan secara *trial*) agar data simulasi yang diperoleh tetap sepola dengan data aslinya. Penetapan nilai toleransi ini dilakukan dengan cara menghitung selisih absolut antara data *Chickpea* yang sudah terurut  $x_{(i)}$  dan data simulasi yang sudah terurut  $\hat{x}_{(i)}$ , yaitu  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon$ . Jika nilai selisih ( $\varepsilon$ ) ini terlalu besar atau terlalu kecil, maka data simulasi akan melenceng dari data aslinya.
5. Lakukan pembangkitan data sebanyak  $l=1.000$  kali, dimana setiap kali pembangkitan data  $l$  diulang sebanyak 3 kali (untuk mendapatkan ukuran sampel  $n=3$ ) dan dilakukan dengan cara sebagai berikut:
  - Untuk perulangan  $i = 1$  atau  $i = 3$ , (untuk mendapatkan data simulasi di sekitar nilai data *Chickpea* yang masih tetap sepola dengan dengan nilai data *Chickpea*), maka
    - (1). Bangkitkan  $\hat{x}_{(i)} \sim N(\bar{x}, s_x)$
    - (2). Hitung toleransi selisih  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon$
    - (3). Ulangi bangkitan pada (1) sampai tercapai kondisi  $\hat{x}_{(1)} \geq x_{(1)}$  dan  $\hat{x}_{(3)} \leq x_{(3)}$  dan  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon = 0,0999$  dapat terpenuhi.
  - Untuk perulangan  $i = 2$ , (diperlakukan berbeda dengan perulangan yang lain dengan tujuan agar tetap diperoleh posisi data tidak berbeda dengan pola data *Chickpea*), maka
    - (1). Bangkitkan  $\hat{x}_{(i)} \sim N(\bar{x}, s_x)$
    - (2). Hitung toleransi selisih  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon$
    - (3). Ulangi bangkitan pada (1) sampai tercapai kondisi  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon = 0,0999$  dapat terpenuhi.

#### **b. Pemodelan Bayesian Unimodal (BUM)**

Untuk membentuk model BUM dari data simulasi dan data *Chickpea*, maka tahapan-tahapan yang dilakukan sebagai berikut:

- Tahap 1. Menetapkan pendekatan distribusi kenormalan untuk setiap ID gen pada data simulasi dan data *Chickpea*.
- Tahap 2. Mengidentifikasi jumlah komponen *mixture* dalam setiap ID gen data simulasi dan data *Chickpea* menggunakan Pengembangan Algoritma RJMCMC.

- Tahap 3. Jika hasil identifikasi Tahap 2 diketahui berdistribusi unimodal, maka selanjutnya membangun 1.000 model normal unimodal pendekatan Bayesian melalui program WinBUGS dengan menerapkan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1). Dalam hal ini, masing-masing parameter yang dibangkitkan sebanyak 5.000 kali. Jika hasil identifikasi Tahap 2 diketahui berdistribusi *mixture*, maka penanganan data selanjutnya menggunakan tahapan pemodelan BMM dan pemodelan BMMA.
- Tahap 4. Melakukan evaluasi pemenuhan asumsi pemodelan Bayesian (*independence, irreducible, aperiodic*, dan *recurrent*) dari model BUM-normal rata-rata yang terbaik berdasarkan hasil Tahap 3 melalui kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diestimasi dalam model.
- Tahap 5. Jika hasil evaluasi model pada Tahap 4 telah memenuhi asumsi yang diharapkan, maka selanjutnya melakukan evaluasi model BUM-normal dengan *goodness of fit* Kolmogorov-Smirnov (KS) sebagaimana yang telah dijabarkan pada sub bab 2.8. Jika hasil Tahap 4 tidak memenuhi asumsi yang diharapkan, maka dilakukan evaluasi terhadap sumber-sumber penyebabnya.
- Tahap 6. Jika hasil evaluasi model pada Tahap 5 menyatakan  $H_0$  tidak ditolak, maka selanjutnya menentukan nilai batas bawah (namakan a) dan batas atas (namakan b) dari HPD dengan *credible interval* 95% melalui penerapan Algoritma 2.3. Jika hasil evaluasi model pada Tahap 5 menyatakan  $H_0$  ditolak, maka model yang dibangun dinyatakan tidak sesuai untuk data dan tidak dilakukan Tahap 7.
- Tahap 7. Menetapkan kelompok data dalam *Down-regulated, Regulated*, dan *Up-regulated* dengan kriteria kelompok data seperti dalam Tabel 3.2.

### c. Pemodelan Bayesian Averaging (BMA)

Untuk membentuk model BMA dari data simulasi dan data *Chickpea*, maka tahapan-tahapan yang dilakukan sebagai berikut:

- Tahap 1. Menetapkan pendekatan distribusi kenormalan untuk setiap ID gen pada data simulasi dan data *Chickpea*.
- Tahap 2. Mengidentifikasi jumlah komponen *mixture* dalam setiap ID gen data simulasi dan data *Chickpea* menggunakan Pengembangan Algoritma RJMCMC.
- Tahap 3. Jika hasil identifikasi Tahap 2 diketahui berdistribusi unimodal, maka selanjutnya membangun 1.000 model normal unimodal pendekatan Bayesian melalui program WinBUGS dengan menerapkan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1). Dalam hal ini, masing-masing parameter yang dibangkitkan sebanyak 5.000 kali. Jika hasil identifikasi Tahap 2 diketahui berdistribusi *mixture*, maka penanganan data selanjutnya menggunakan tahapan pemodelan BMM dan pemodelan BMMA.
- Tahap 4. Melakukan evaluasi pemenuhan asumsi pemodelan Bayesian (*independence*, *irreducible*, *aperiodic*, dan *recurrent*) dari model BUM-normal rata-rata yang terbaik berdasarkan hasil Tahap 3 melalui kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diestimasi dalam model. Jika hasil Tahap 4 ini tidak memenuhi asumsi yang diharapkan, maka dilakukan evaluasi terhadap sumber-sumber penyebabnya.
- Tahap 5. Jika hasil evaluasi model pada Tahap 4 telah memenuhi asumsi yang diharapkan, maka selanjutnya melakukan seleksi model dengan metode seleksi Occam's Window sebagaimana yang telah dijabarkan pada sub bab 2.6.2. dengan menerapkan persamaan (2.39).
- Tahap 6. Membentuk model BMA-normal berdasarkan hasil seleksi model Tahap 5 dan selanjutnya melakukan evaluasi model dengan kriteria *goodness of fit* KS sebagaimana yang telah dijabarkan pada sub bab 2.8.
- Tahap 7. Jika hasil evaluasi model dengan kriteria *goodness of fit* KS pada Tahap 6 menyatakan  $H_0$  tidak ditolak, maka selanjutnya menentukan nilai batas bawah (namakan a) dan batas atas (namakan b) dari HPD dengan *credible interval* 95% melalui penerapan Algoritma 2.3. Jika hasil evaluasi model pada Tahap 6 menyatakan  $H_0$  ditolak, maka model

yang dibangun dinyatakan tidak sesuai untuk data dan tidak dilakukan Tahap 8.

Tahap 8. Menetapkan kelompok data dalam *Down-regulated*, *Regulated*, dan *Up-regulated* dengan kriteria kelompok data seperti dalam Tabel 3.2.

#### **d. Pemodelan Bayesian *Mixture* (BMM)**

Untuk membentuk model BMM dari data simulasi dan data *Chickpea*, maka tahapan-tahapan yang dilakukan sebagai berikut:

Tahap 1. Menetapkan pendekatan distribusi kenormalan untuk setiap ID gen pada data simulasi dan data *Chickpea*.

Tahap 2. Mengidentifikasi jumlah komponen *mixture* dalam setiap ID gen data simulasi dan data *Chickpea* menggunakan Pengembangan Algoritma RJMCMC.

Tahap 3. Jika hasil identifikasi Tahap 2 diketahui berdistribusi *mixture* dengan dua dan atau tiga komponen, maka selanjutnya membangun 1.000 model *mixture* normal dua dan tiga komponen pendekatan Bayesian melalui program WinBUGS dengan menerapkan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1). Dalam hal ini, masing-masing parameter yang dibangkitkan sebanyak 5.000 kali. Jika hasil identifikasi Tahap 2 ini diketahui berdistribusi bukan *mixture* (unimodal), maka penanganan data selanjutnya menggunakan tahapan pemodelan BUM dan pemodelan BMA.

Tahap 4. Melakukan evaluasi pemenuhan asumsi pemodelan Bayesian (*independence*, *irreducible*, *aperiodic*, dan *recurrent*) dari model BMM-normal rata-rata yang terbaik berdasarkan hasil Tahap 3 melalui kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diestimasi dalam model.

Tahap 5. Jika hasil evaluasi model pada Tahap 4 telah memenuhi asumsi yang diharapkan, maka selanjutnya melakukan evaluasi model BMM-normal dua dan atau tiga komponen dengan *goodness of fit* Kolmogorov-Smirnov (KS) sebagaimana yang telah dijabarkan pada sub bab 2.8.



Jika hasil Tahap 4 tidak memenuhi asumsi yang diharapkan, maka dilakukan evaluasi terhadap sumber-sumber penyebabnya.

- Tahap 6. Jika hasil evaluasi model pada Tahap 5 menyatakan  $H_0$  tidak ditolak, maka selanjutnya menentukan nilai batas bawah (namakan a) dan batas atas (namakan b) dari HPD dengan *credible interval* 95% melalui penerapan Algoritma 2.3. Jika hasil evaluasi model pada Tahap 5 menyatakan  $H_0$  ditolak, maka model yang dibangun dinyatakan tidak sesuai untuk data dan tidak dilakukan penentuan nilai-nilai batas HPD dengan *credible interval* 95% serta tidak dilakukan Tahap 7.
- Tahap 7. Menetapkan kelompok data dalam *Down-regulated*, *Regulated*, dan *Up-regulated* dengan kriteria kelompok data seperti dalam Tabel 3.2.

**e. Pemodelan Bayesian *Mixture Averaging* (BMMA)**

Untuk membentuk model BMMA dari data simulasi dan data *Chickpea* maka tahapan-tahapan yang dilakukan sebagai berikut:

- Tahap 1. Menetapkan pendekatan distribusi kenormalan untuk setiap ID gen pada data simulasi dan data *Chickpea*.
- Tahap 2. Mengidentifikasi jumlah komponen *mixture* dalam setiap ID gen data simulasi dan data *Chickpea* menggunakan Pengembangan Algoritma RJMCMC.
- Tahap 3. Jika hasil identifikasi Tahap 2 diketahui berdistribusi *mixture* dengan dua dan atau tiga komponen, maka selanjutnya membangun 1.000 model *mixture* normal dua dan tiga komponen pendekatan Bayesian melalui program WinBUGS dengan menerapkan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1). Dalam hal ini, masing-masing parameter yang dibangkitkan sebanyak 5.000 kali. Jika hasil identifikasi Tahap 2 ini diketahui berdistribusi bukan *mixture* (unimodal), maka penanganan data selanjutnya menggunakan tahapan pemodelan BUM dan pemodelan BMA.
- Tahap 4. Melakukan evaluasi pemenuhan asumsi pemodelan Bayesian (*independence*, *irreducible*, *aperiodic*, dan *recurrent*) dari model BMM-normal rata-rata yang terbaik berdasarkan hasil Tahap 3 melalui

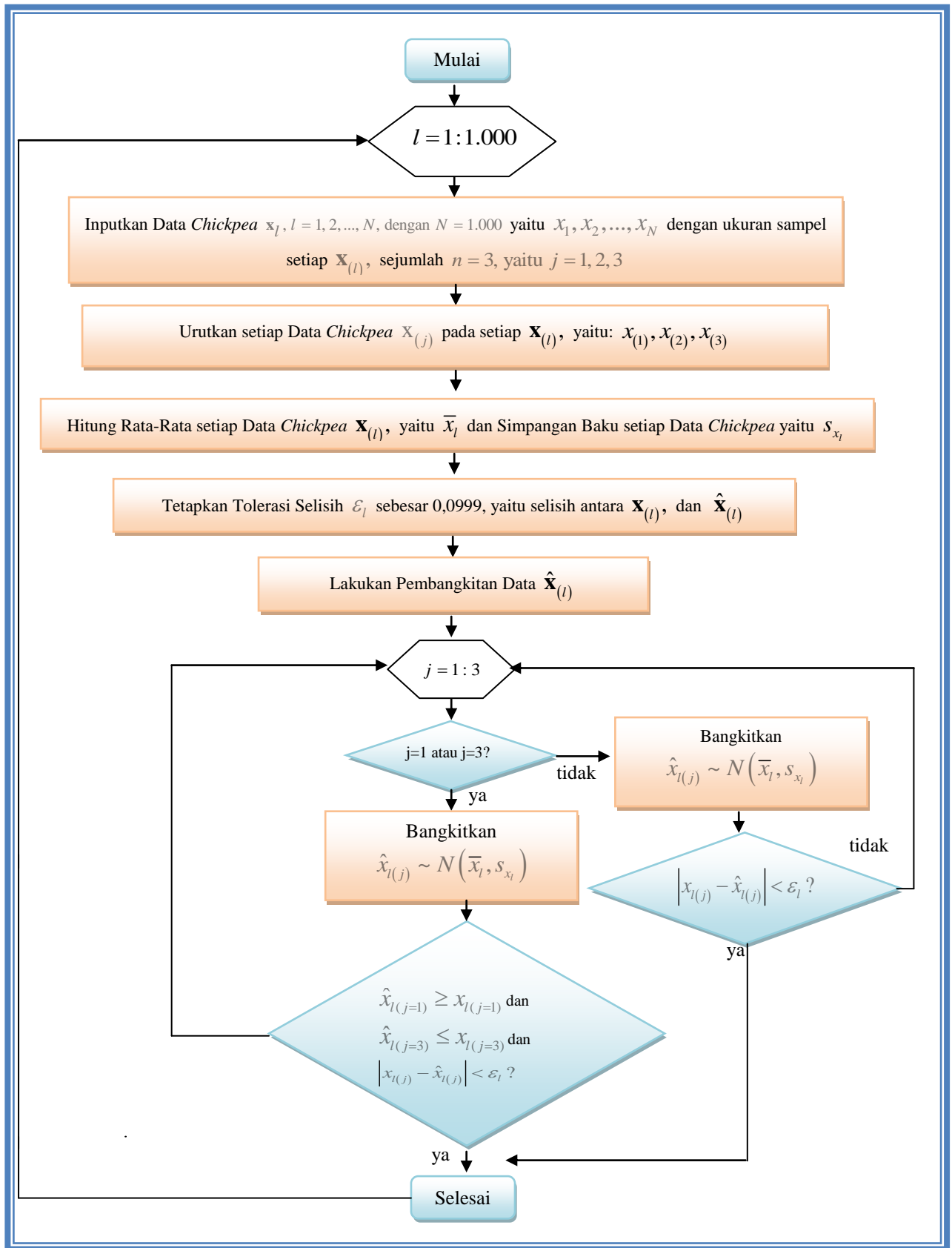
kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diestimasi dalam model.

- Tahap 5. Jika hasil evaluasi model pada Tahap 4 telah memenuhi asumsi yang diharapkan, maka selanjutnya melakukan seleksi model dengan metode seleksi Occam's Window sebagaimana yang telah dijabarkan dalam sub bab 2.6.2 dengan menerapkan persamaan (2.39). Jika hasil Tahap 4 tidak memenuhi asumsi yang diharapkan, maka dilakukan evaluasi terhadap sumber-sumber penyebabnya.
- Tahap 6. Membentuk model BMMA-normal berdasarkan hasil seleksi model Tahap 5 dan selanjutnya melakukan evaluasi model dengan kriteria *goodness of fit* KS sebagaimana yang telah dijabarkan pada sub bab 2.8.
- Tahap 7. Jika hasil evaluasi model pada Tahap 6 menyatakan  $H_0$  tidak ditolak, maka selanjutnya menentukan nilai batas bawah (namakan a) dan batas atas (namakan b) dari HPD dengan *credible interval* 95% melalui penerapan Algoritma 2.3. Jika hasil evaluasi model pada Tahap 6 menyatakan  $H_0$  ditolak, maka model yang dibangun dinyatakan tidak sesuai untuk data dan tidak dilakukan penentuan nilai-nilai batas HPD dengan *credible interval* 95% serta tidak dilakukan Tahap 8.
- Tahap 8. Menetapkan kelompok data dalam *Down-regulated*, *Regulated*, dan *Up-regulated* dengan kriteria kelompok data seperti dalam Tabel 3.2.

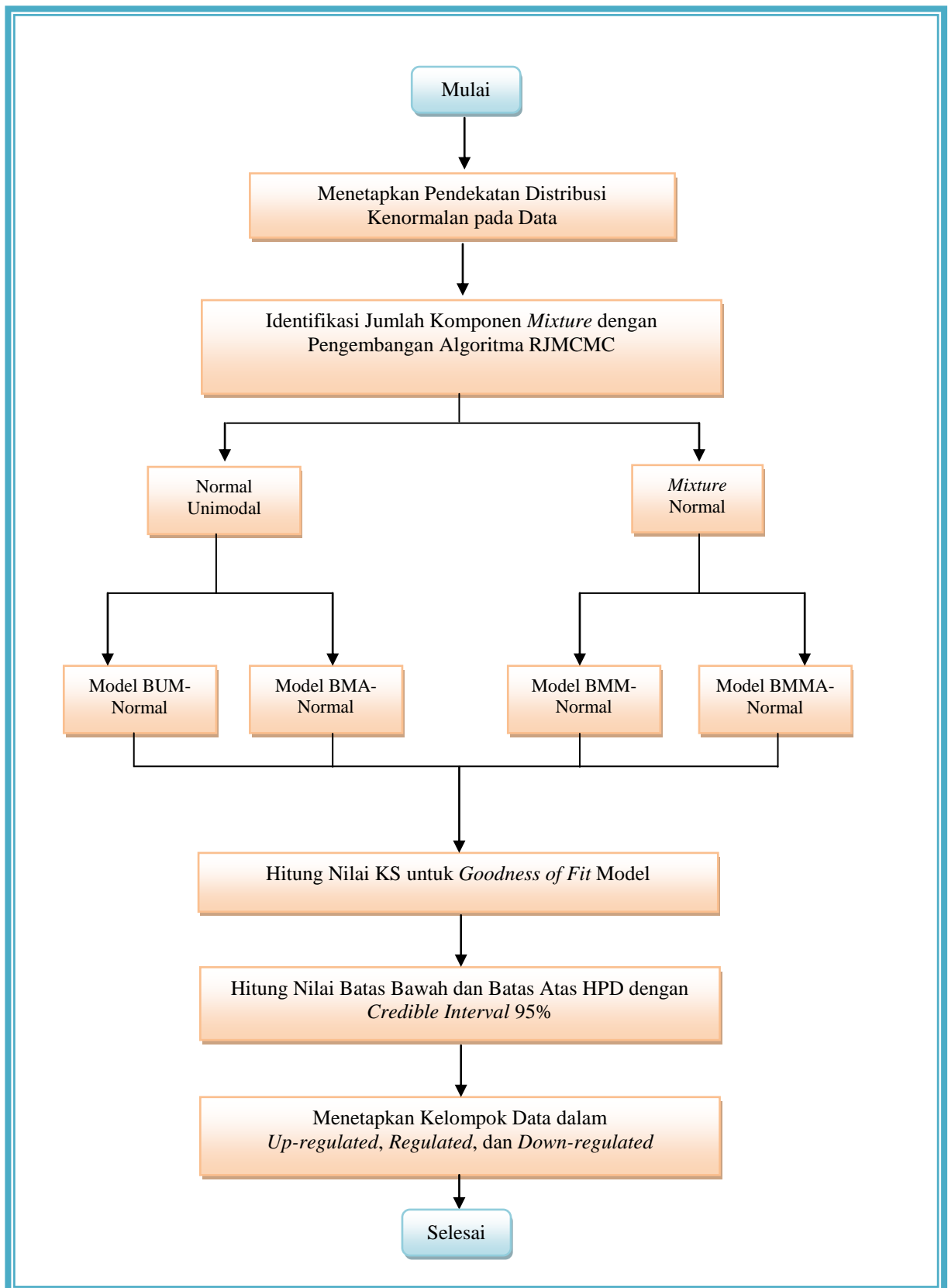
Metode penelitian yang telah dijabarkan akan dibuat bagan alir metode penelitian beserta penjabarannya sebagaimana disajikan pada Gambar 3.3 sampai dengan Gambar 3.6.

**Tabel 3.2.** Kriteria Kelompok ID Gen berdasarkan nilai HPD dalam Pemodelan

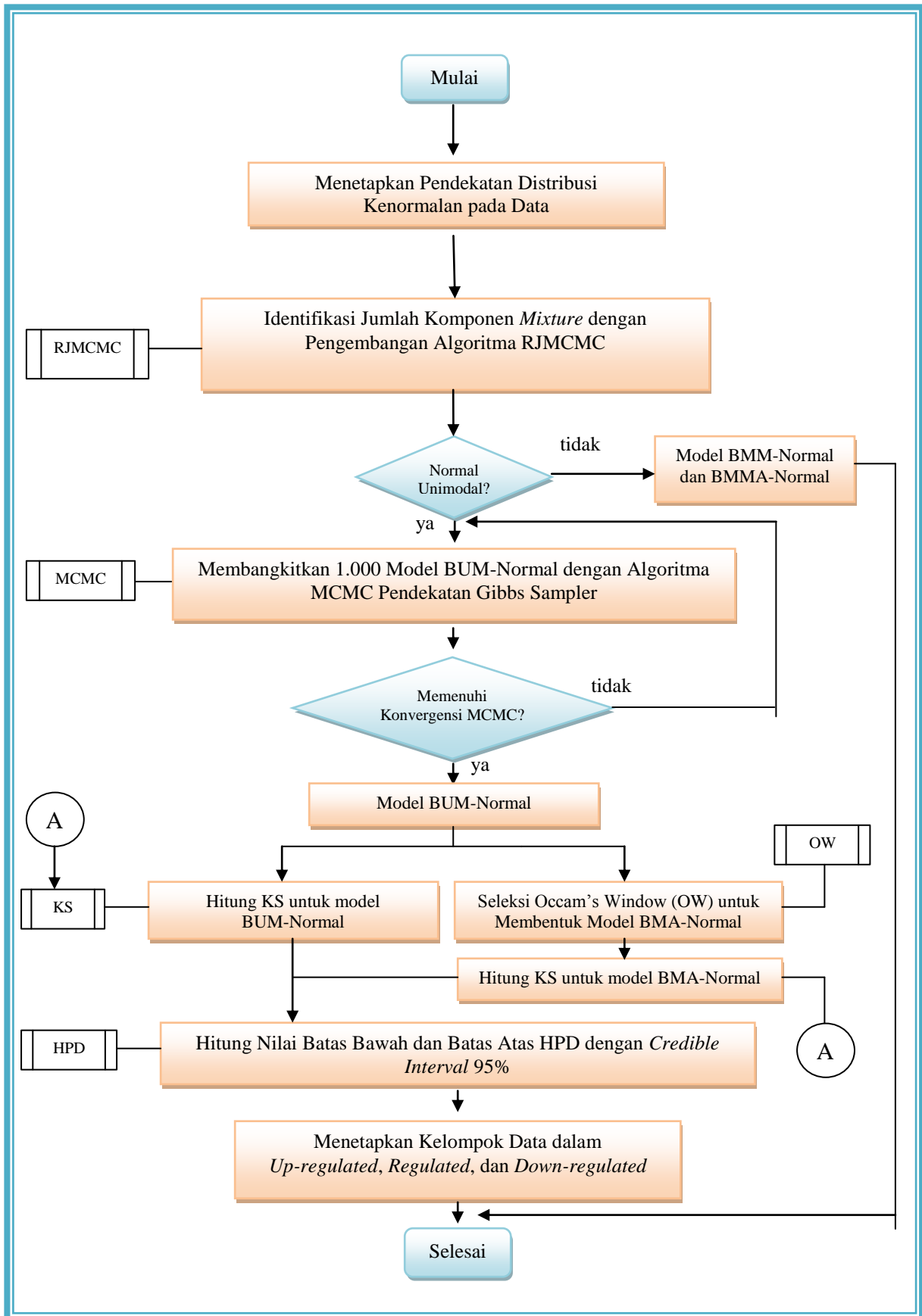
No	Kelompok ID Gen	Kriteria Menurut Nilai HPD	Deskripsi Data
1	<i>Down-regulated</i>	Batas Bawah (namakan a): (+) Batas Atas (namakan b): (+)  (ID gen berada di antara kedua nilai batas yang bernilai positif)	ID gen pada kelompok ini adalah ID gen yang tahan dari serangan penyakit <i>Ascochyta Rabiei</i> dan diharapkan dapat meningkatkan produksi Kacang <i>Chickpea</i>
2	<i>Regulated</i>	Batas Bawah (namakan a): (-) Batas Atas (namakan b): (+)  (ID gen berada di antara kedua nilai batas yang bernilai negatif dan positif)	ID gen pada kelompok ini adalah ID gen yang tidak cukup tahan dari serangan penyakit <i>Ascochyta Rabiei</i> sehingga kurang mampu meningkatkan produksi Kacang <i>Chickpea</i>
3	<i>Up-regulated</i>	Batas Bawah (namakan a): (-) Batas Atas (namakan b): (-)  (ID gen berada di antara kedua nilai batas yang bernilai negatif)	ID gen pada kelompok ini adalah ID gen yang tidak tahan terhadap serangan penyakit <i>Ascochyta Rabiei</i> dan dapat menurunkan produksi Kacang <i>Chickpea</i>



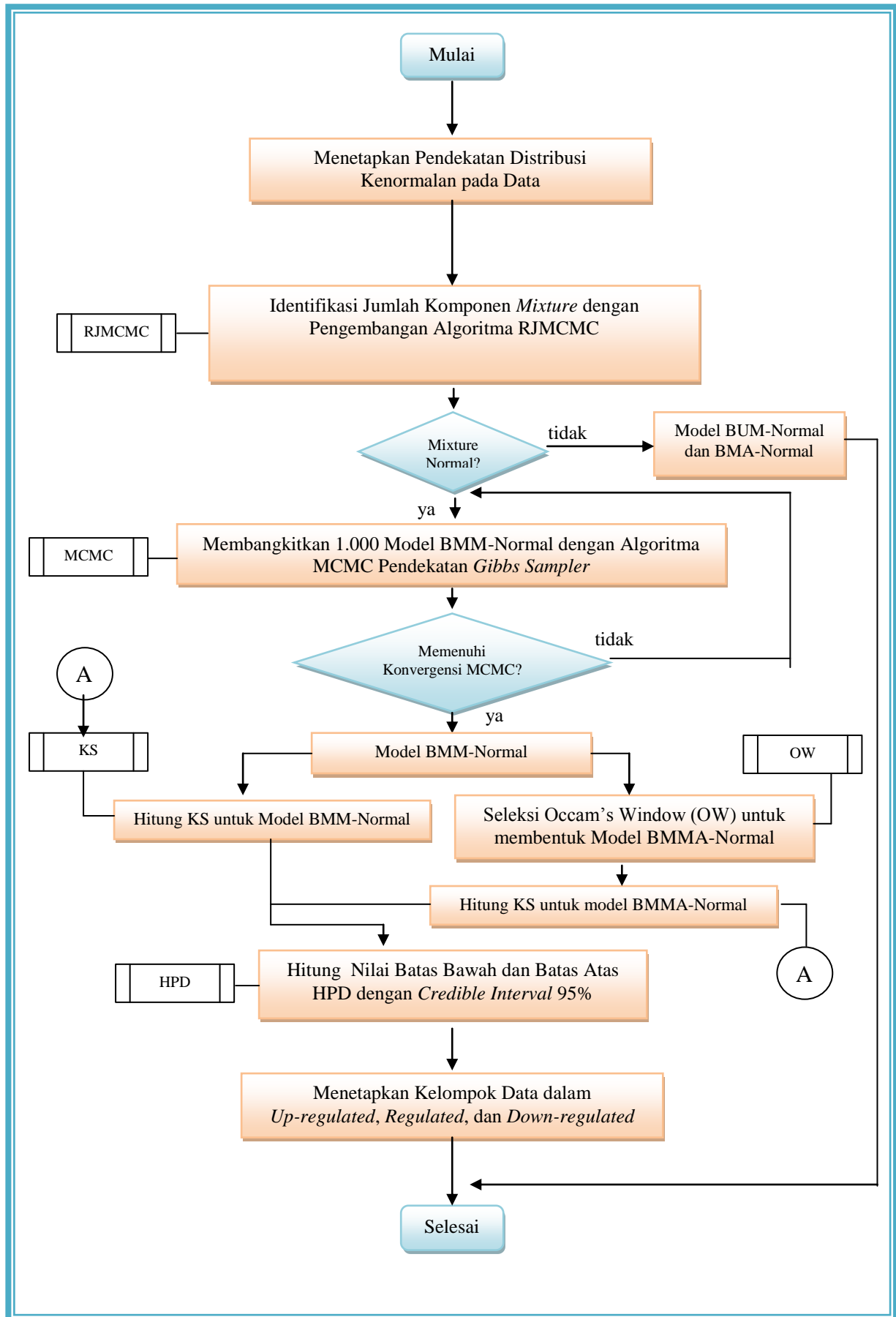
**Gambar 3.3.** Bagan Alir Metode Penelitian untuk Pembangkitan Data Simulasi



**Gambar 3.4.** Bagan Alir Metode Penelitian Secara Umum untuk Pemodelan



**Gambar 3.5.** Bagan Alir Metode Penelitian untuk Pemodelan BUM-Normal dan Pemodelan BMA-Normal Unimodal



**Gambar 3.6.** Bagan Alir Metode Penelitian untuk Pemodelan BMM-Normal dan Pemodelan BMMA-Normal

## BAB IV

### PENGEMBANGAN ALGORITMA RJMCMC DAN SKENARIO PEMBANGKITAN DATA SIMULASI UNTUK DATA *CHICKPEA* DI INDONESIA

Dalam bab ini akan disajikan Algoritma MCMC dan pengembangannya, skenario pembangkitan data simulasi melalui algoritma pembangkitan data simulasi, dan validasi model dengan *goodness of fit* Kolmogorov-Smirnov (KS). Implementasi dari semua algoritma yang dibangun dilakukan pada data simulasi dan data *Chickpea*.

#### 4.1. Algoritma RJMCMC

Algoritma RJMCMC digunakan untuk mengetahui berapa jumlah komponen *mixture* yang tidak diketahui dalam data berdasarkan konsep data *driven*. Algoritma RJMCMC telah banyak digunakan pada berbagai kasus, salah satunya yang digunakan oleh Feng (2013) yang disebut sebagai Algoritma 4.1. Algoritma ini diterapkan pada kasus data Feng yang memiliki distribusi *mixture* normal. Hasil identifikasinya menunjukkan sangat baik dalam menetapkan jumlah komponen *mixture* dalam data Feng tersebut. Akan tetapi pada saat Algoritma 4.1 diterapkan pada data *Chickpea* yang memiliki ciri distribusi yang sama yaitu normal, Algoritma 4.1 hanya mampu mengidentifikasi data yang berdistribusi *mixture* normal saja dan tidak mampu mengidentifikasi yang berdistribusi normal unimodal (bukan *mixture*). Oleh karena itu perlu pengembangan Algoritma 4.1 agar diperoleh suatu algoritma baru yang mampu mengidentifikasi distribusi *mixture* normal dan sekaligus distribusi normal unimodal (bukan *mixture*) dalam data.

Pengembangan dari Algoritma 4.1 yang disebut sebagai Algoritma 4.4, yaitu Pengembangan Algoritma RJMCMC ini merupakan algoritma gabungan dari Algoritma RJMCMC, Algoritma Proses Pendahuluan (*Preliminary Process Algorithm*) dan enam langkah pertama dari Algoritma Proses Pemulusan (*Smoothing Process Algorithm*). Algoritma Proses Pendahuluan dan Algoritma Proses Pemulusan ditemukan oleh Astuti dkk. (2015c) dan digunakan untuk



mendeteksi jumlah komponen *mixture* dalam data. Kedua algoritma ini dibangun berdasarkan pembangkitan data simulasi yang berciri data *Chickpea* di Indonesia dengan menggunakan konsep jarak antar titik data. Melalui Pengembangan Algoritma RJMCMC dapat diidentifikasi jumlah komponen *mixture* dalam data *Chickpea* yang berdistribusi multimodal (bimodal dan trimodal) dan sekaligus yang berdistribusi unimodal.

Untuk mengimplementasi Algoritma 4.1 didasarkan atas konsep MCMC sebagaimana yang dituangkan dalam Algoritma 2.1. Langkah pertama diawali dengan set data  $\mathbf{x}$  dalam bentuk  $c(x_1, x_2, \dots, x_n)$  dengan ukuran sampel data sejumlah  $n$  sebagai panjang dari data  $\mathbf{x}$  dan dinotasikan dengan  $length(\mathbf{x})$ . Langkah selanjutnya adalah menentukan inisialisasi untuk parameter-parameter model sesuai dengan distribusi *mixture* dari data  $\mathbf{x}$  dan menetapkan iterasi *sweep* sebanyak minimal 200.000. Dalam hal ini, jika distribusi dari data  $\mathbf{x}$  adalah distribusi *mixture* normal, maka inisialisasi dilakukan untuk parameter  $\mathbf{w}$ ,  $\mu$ ,  $\sigma^2$ , dan  $\mathbf{z}$  dengan fungsi *mixture* normalnya dituliskan sebagai berikut:  $(w[i] * dnorm(\mathbf{x}, \mu[i], \sqrt{\sigma^2[i]}))$ . Selanjutnya adalah menjalankan program RJMCMC dengan menggunakan gabungan dari konsep *Metropolis Hastings* yang telah dituangkan dalam Algoritma 2.2 untuk mengatasi jumlah komponen *mixture* yang tidak/belum diketahui dan konsep *Reversible Jump* yang menggunakan prinsip *bird-death*, *split-merge* dengan enam tipe pergerakan yaitu (1) *Updating w*, (2) *Updating  $\theta$*  dalam hal ini  $\mu$ ,  $\sigma^2$ , (3) *Updating z*, (4) *Updating Hiperparameter  $\beta$* , (5) *Split/Merge* komponen-komponen, dan (6) *Birth/Death* dari komponen kosong. Untuk langkah-langkah *split-merge*, *bird-death* dalam konsep *Reversible Jump* sebagaimana telah dijelaskan dalam sub bab 2.4.7.

Secara lengkap, langkah-langkah dari Algoritma RJMCMC yang dibangun oleh Feng (2013) dituangkan dalam Algoritma 4.1.

Langkah 1. Lakukan set data  $\mathbf{x}$  yang akan dianalisis sebagai  $c(x_1, x_2, \dots, x_n)$  dengan informasi ukuran sampel  $n$  sebagai  $length(\mathbf{x})$  dan nilai maksimum dari data  $\mathbf{x}$  sebagai  $\max \mathbf{x}$ .

- Langkah 2. Menentukan inisialisasi untuk nilai  $\mathbf{w}$ ,  $\mu$ ,  $\sigma^2$ , dan  $\mathbf{z}$  yang dibangun dengan fungsi *mixture* normal  $(w[i] * dnorm(\mathbf{x}, mu[i], sqrt(sigma2[i])))$  serta iterasi nilai *sweep* sebanyak 200.000.
- Langkah 3. Jalankan program RJMCMC dengan prinsip *bird-death*, *split-merge* untuk data  $\mathbf{x}$  berdasarkan inisialisasi  $\mathbf{w}$ ,  $\mu$ ,  $\sigma^2$ ,  $\mathbf{z}$ , fungsi data  $\mathbf{x}$  dan  $\max \mathbf{x}$ . Dalam menjalankan program RJMCMC, sebagai inputnya adalah data  $\mathbf{x}$ , jumlah iterasi (jumlah *sweep*=200.000),  $k_{\max} = n$ , mulai  $k=1$ , nilai  $\mathbf{w}$ ,  $\mu$ ,  $\sigma^2$ ,  $\mathbf{z}$  dan fungsi *mixture* normal.
- Langkah 4. Simpan hasil jumlah komponen *mixture* dari densitas normal yang berupa bilangan bulat (diskrit) dengan operasi matematika *round* di setiap iterasi (*sweep*).
- Langkah 5. Hitung jumlah komponen *mixture* dari densitas normal data  $\mathbf{x}$  dengan rumus  $round\left(\frac{\text{jumlah komponen mixture setiap iterasi}}{\text{banyaknya iterasi}}\right)$  dimana banyaknya iterasi *sweep* sebesar 200.000.

#### 4.2. Algoritma Proses Pendahuluan (*Preliminary Process Algorithm*)

Algoritma Proses Pendahuluan (*Preliminary Process Algorithm*) (Algoritma 4.2) dibangun untuk menentukan nilai-nilai batas sebagai penentu apakah distribusi data unimodal (bukan *mixture*) ataukah multimodal (*mixture*). Dalam hal ini telah ditetapkan ada lima langkah dari Algoritma 4.1 dan ditemukan

dua nilai batas yang disebut sebagai  $c_{\text{batas1}} = \left[ \frac{(\bar{X})_{\text{total}}}{2} + 0,017485114 \right]$  dan

$d_{\text{batas2}} = \left[ \frac{(\bar{X})_{\text{total}}}{3} - 0,021175691 \right]$  yang dihitung berdasarkan rata-rata total dari

seluruh ID gen yang diamati dengan penambahan dan pengurangan suatu konstanta tertentu sehingga berdasarkan nilai batas ini dapat diketahui apakah kisaran (*range*) data lebih besar atau lebih kecil dari nilai batas yang telah

ditetapkan. Kedua nilai batas ini ditetapkan berdasarkan hasil dari pembangkitan data simulasi sebanyak 100.000 kali.

Tahapan-tahapan dari Algoritma 4.2 dapat dijabarkan sebagai berikut:

**Algoritma 4.2.** Algoritma Proses Pendahuluan (*Preliminary Process Algorithm*)

Langkah 1. Urutkan data pada masing-masing ID gen, dari terkecil sampai terbesar, yaitu  $X_{(1)p}$ ,  $X_{(2)p}$ ,  $X_{(3)p}$  dengan  $p = 1, 2, 3, \dots, q$  dan  $q$  adalah jumlah ID gen yang diamati.

Langkah 2. Hitung rata-rata data pada setiap ID gen berdasarkan hasil pada

Langkah 1 dengan rumus  $\bar{X}_p = \frac{\sum_{i=1}^3 X_{(i)p}}{3}$ .

Langkah 3. Hitung rata-rata total untuk seluruh ID gen yang diamati berdasarkan hasil pada Langkah 2 dengan rumus

$$(\bar{X})_{\text{total}} = \frac{\sum_{j=1}^q \bar{X}_p}{q}.$$

Langkah 4. Hitung nilai batas pertama dengan rumus

$$c_{\text{batas1}} = \left[ \frac{\bar{X}_{\text{total}}}{2} + 0,017485114 \right].$$

Nilai batas ini merupakan hasil temuan dari studi simulasi Astuti dkk. (2015c).

Langkah 5. Hitung nilai batas kedua dengan rumus

$$d_{\text{batas2}} = \left[ \frac{\bar{X}_{\text{total}}}{3} - 0,021175691 \right].$$

Nilai batas ini merupakan hasil temuan dari penulis melalui studi simulasi yang dilakukan (Astuti dkk., 2015c).

**4.3. Algoritma Proses Pemulusan (*Smoothing Process Algorithm*)**

Algoritma Proses Pemulusan (Algoritma 4.3) digunakan untuk proses pengoptimuman hasil identifikasi jumlah komponen *mixture* dalam data dengan menggunakan konsep jarak antar titik data. Dengan Algoritma 4.3 dapat ditentukan distribusi unimodal, distribusi *mixture* dan sekaligus kemencengan dari distribusi *mixture*. Khusus untuk menentukan distribusi data unimodal, nilai

kisaran (*range*) dari data yang disimbolkan dengan  $c_p$  harus lebih kecil atau sama dari nilai batas  $c_{batas1}$ . Penggunaan Algoritma 4.3 ini masih terbatas pada penentuan jumlah komponen *mixture* maksimum sebesar tiga. Hal ini dikarenakan data simulasi yang dibangkitkan berdasarkan data *Chickpea* kasus di Indonesia hanya memiliki ukuran sampel sebesar tiga ( $n=3$ ) untuk setiap ID gen. Selain itu, rumitnya proses penetapan jumlah komponen *mixture* ini menjadi kendala yang cukup penting apabila diterapkan untuk kasus data yang memiliki jumlah komponen *mixture* lebih dari tiga.

Tahapan-tahapan dari Algoritma 4.3 yang merupakan hasil temuan penulis (Astuti dkk., 2015c) melalui studi simulasi, dapat dijabarkan berikut:

**Algoritma 4.3.** Algoritma Proses Pemulusan (*Smoothing Process Algorithm*)

Langkah 1. Lakukan Langkah 1 sampai dengan Langkah 5 dari Algoritma 4.2.

Langkah 2. Hitung nilai simpangan baku pada masing-masing ID gen ke- $p$  dari  $X_{(1)p}$  dan  $X_{(2)p}$  yaitu  $stdev_{12p}$  yang dihitung dengan rumus

$$stdev_{12p} = \frac{\sum_{i=1}^2 (X_{(i)p} - \bar{X}_{(1)(2)p})^2}{(2-1)} \quad \text{dan simpangan baku untuk}$$

masing-masing ID gen ke- $p$  dari  $X_{(2)p}$  dan  $X_{(3)p}$  yaitu  $stdev_{23p}$

$$\text{yang dihitung dengan rumus } stdev_{23p} = \frac{\sum_{i=2}^3 (X_{(i)p} - \bar{X}_{(2)(3)p})^2}{(2-1)}.$$

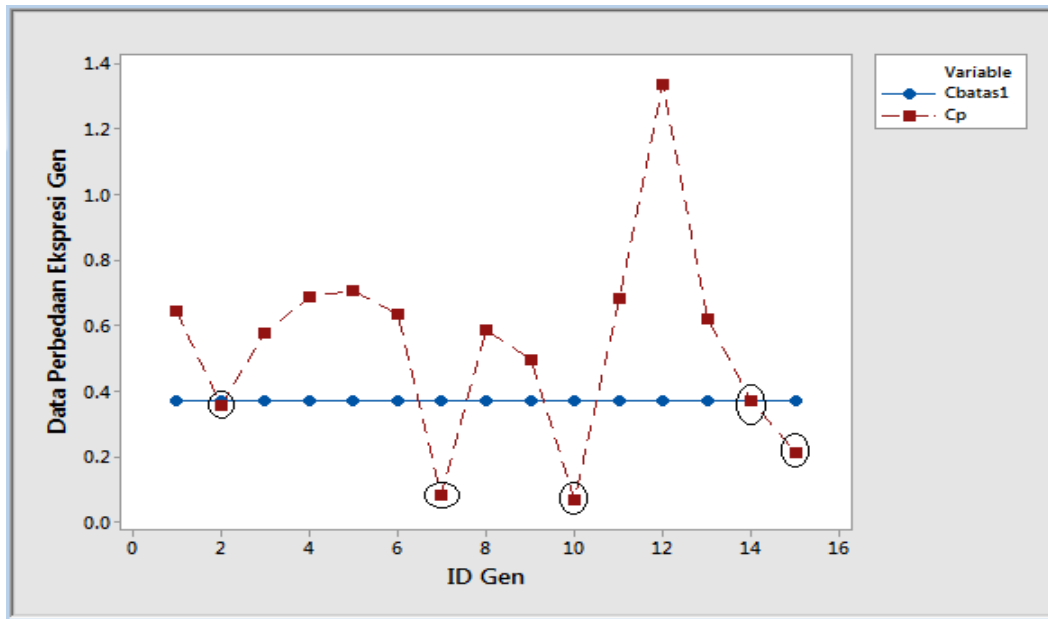
Langkah 3. Hitung nilai minimum dari simpangan baku untuk masing-masing ID gen ke- $p$  yang didasarkan pada hasil Langkah 2 yaitustdev<sub>min- $p$</sub>  = min (stdev<sub>12 $p$</sub> , stdev<sub>23 $p$</sub> ).

Langkah 4. Hitung nilai batas bawah dan batas atas dari data  $X_{(i)}$  untuk masing ID gen ke- $p$  dengan rumus  $X_{(i)p} \pm (1,96 \times stdev_{min-p} / \sqrt{3})$ . Batas bawah dan batas atas untuk  $X_{(i)p}$  disebut  $lwX_{(i)p}$  and  $upX_{(i)p}$ .

Langkah 5. Hitung nilai dari  $a_{1p} = abs(X_{(1)p} - X_{(2)p})$ ,  $a_{2p} = abs(X_{(2)p} - X_{(3)p})$ ,  $c_p = abs(X_{(1)p} - X_{(3)p})$  dan  $d_p = abs(a_{1p} - X_{(3)p})$   $d_p = abs(a_{1p} - a_{2p})$  untuk masing-masing ID gen ke- $p$ .

- Langkah 6. Identifikasi apakah nilai dari  $c_p < c_{\text{batas1}}$ . Jika benar maka distribusi ID gen ke- $p$  adalah unimodal.
- Langkah 7. Identifikasi apakah nilai dari  $c_p \geq c_{\text{batas1}}$  dan  $d_p \geq d_{\text{batas2}}$  dan  $upX_{(1)p} < lwX_{(1)p}$  dan  $lwX_{(3)p} \leq upX_{(2)p}$ . Jika benar, maka data tersebut miring ke kiri dan membentuk distribusi *mixture* dengan dua komponen.
- Langkah 8. Identifikasi apakah nilai dari  $c_p \geq c_{\text{batas1}}$  dan  $d_p \geq d_{\text{batas2}}$  dan  $lwX_{(2)p} \leq upX_{(1)p}$  dan  $upX_{(2)p} < lwX_{(3)p}$ . Jika benar, maka data tersebut miring ke kanan dan membentuk distribusi *mixture* dengan dua komponen.
- Langkah 9. Identifikasi apakah nilai dari  $c_p \geq c_{\text{batas1}}$  dan  $d_p < d_{\text{batas2}}$ . Jika benar, maka data tersebut membentuk distribusi *mixture* dengan tiga komponen.
- Langkah 10. Identifikasi jumlah ID gen yang tidak dapat diidentifikasi.

Ilustrasi dari konsep Algoritma 4.2 dan enam langkah pertama dari Algoritma 4.3 (khusus untuk menentukan distribusi data unimodal) dapat digambarkan dalam Gambar 4.1 berikut:



**Gambar 4.1.** Ilustrasi Konsep Algoritma 4.1 dan Enam Langkah Pertama Algoritma 4.2 untuk Identifikasi Distribusi Data Unimodal

Berdasarkan ilustrasi dalam Gambar 4.1 dapat ditunjukkan ada lima ID gen yang memiliki distribusi unimodal yaitu ID gen 2, ID gen 7, ID gen 10, ID gen 14 dan ID gen 15 karena lima ID gen ini memiliki nilai  $c_p \leq c_{\text{batas1}}$ .

Akurasi dari kedua algoritma tersebut telah dipublikasikan penulis (Astuti dkk., 2015c), yaitu Algoritma 4.2 dan Algoritma 4.3 menunjukkan hasil yang sangat baik yaitu sebesar 99,369%. Artinya dari 100.000 data simulasi yang dibangkitkan ada sebanyak 99.360 data dapat diidentifikasi dengan baik dan tepat jumlah komponen *mixture*-nya yang sesuai dengan kondisi data *Chickpea*. Sedangkan penerapan kedua algoritma ini terhadap data *microarray* Do dkk. (2004) menunjukkan hasil akurasi sebesar 99,9094%. Artinya dari 100.000 data simulasi yang dibangkitkan ada sebanyak 99.909 data dapat diidentifikasi dengan baik dan tepat jumlah komponen *mixture*-nya yang sesuai dengan kondisi data Do dkk. (Astuti dkk., 2015c).

#### **4.4. Pengembangan Algoritma RJMCMC**

Sebagaimana yang telah disebutkan sebelumnya bahwa Algoritma 4.1 memiliki kendala dalam mengidentifikasi distribusi normal unimodal (bukan *mixture*) dalam data *Chickpea* sehingga Algoritma 4.1 perlu dikembangkan. Pengembangan dari Algoritma 4.1 yang disebut sebagai Pengembangan Algoritma RJMCMC (Algoritma 4.4), dibangun dengan menggabungkan Algoritma 4.1, Algoritma 4.2 dan enam langkah pertama dari Algoritma 4.3. Melalui Algoritma 4.4 ini dapat diidentifikasi jumlah komponen *mixture* dalam data *Chickpea* yang berdistribusi normal multimodal (bimodal dan trimodal) dan sekaligus yang berdistribusi normal unimodal.

Penggunaan Algoritma 4.4 untuk menentukan jumlah komponen *mixture* dalam data *Chickpea* lebih mudah, *simple*, fleksibel dan telah terbukti cukup handal hasil identifikasinya pada berbagai kasus data bahkan bisa untuk berbagai jumlah komponen *mixture* dalam data (lebih dari tiga komponen *mixture*) dibandingkan dengan penggunaan Algoritma 4.2. Oleh karena itu pengembangan Algoritma RJMCMC ini menggabungkan beberapa keunggulan dari Algoritma 4.1, Algoritma 4.2 dan Algoritma 4.3, sehingga secara menyeluruh mampu

mengidentifikasi jumlah komponen *mixture* dalam data *Chickpea* di Indonesia, baik yang berdistribusi unimodal maupun yang multimodal.

Tahapan-tahapan secara lengkap dari Algoritma 4.4, dimana Langkah 1 sampai dengan Langkah 10 merupakan hasil temuan studi simulasi dari penulis (Astuti dkk., 2015c) dan Langkah 11 sampai dengan Langkah 15 merupakan hasil temuan Feng (2013), disajikan berikut ini.

**Algoritma 4.4.** Pengembangan Algoritma RJMCMC

Langkah 1. Urutkan data pada masing-masing ID gen, dari terkecil sampai terbesar, yaitu  $X_{(1)p}$ ,  $X_{(2)p}$ ,  $X_{(3)p}$  dengan  $p = 1, 2, 3, \dots, q$  dan  $q$  adalah jumlah ID gen yang diamati.

Langkah 2. Hitung rata-rata data pada setiap ID gen berdasarkan hasil pada

Langkah 1 dengan rumus  $\bar{X}_p = \frac{\sum_{i=1}^3 X_{(i)p}}{3}$ .

Langkah 3. Hitung rata-rata total untuk seluruh ID gen yang diamati berdasarkan hasil pada Langkah 2 dengan rumus

$$(\bar{X})_{\text{total}} = \frac{\sum_{j=1}^q \bar{X}_p}{q}.$$

Langkah 4. nilai batas pertama dengan rumus

$$c_{\text{batas1}} = \left[ \frac{(\bar{X})_{\text{total}}}{2} + 0,017485114 \right].$$

Langkah 5. Hitung nilai batas kedua dengan rumus

$$d_{\text{batas2}} = \left[ \frac{(\bar{X})_{\text{total}}}{3} - 0,021175691 \right].$$

Langkah 6. Hitung nilai simpangan baku pada masing-masing ID gen ke- $p$  dari  $X_{(1)p}$  dan  $X_{(2)p}$  yaitu  $\text{stdev}_{12p}$  yang dihitung dengan rumus

$$\text{stdev}_{12p} = \frac{\sum_{i=1}^2 (X_{(i)p} - \bar{X}_{(1)(2)p})^2}{(2-1)} \quad \text{dan simpangan baku untuk}$$

masing-masing ID gen ke- $p$  dari  $X_{(2)p}$  dan  $X_{(3)p}$  yaitu  $\text{stdev}_{23p}$

$$\text{yang dihitung dengan rumus } \text{stdev}_{23p} = \frac{\sum_{i=2}^3 (X_{(i)p} - \bar{X}_{(2)(3)p})^2}{(2-1)}.$$

Langkah 7. Hitung nilai minimum dari simpangan baku untuk masing-masing ID gen ke- $p$  yang didasarkan pada hasil Langkah 7 yaitu  $\text{stdev}_{\min-p} = \min(\text{stdev}_{12p}, \text{stdev}_{23p})$ .

Langkah 8. Hitung nilai batas bawah dan batas atas dari data  $X_{(i)}$  untuk masing ID gen ke- $p$  dengan rumus  $X_{(i)p} \pm (1,96 \times \text{stdev}_{\min-p} / \sqrt{3})$ . Batas bawah dan batas atas untuk  $X_{(i)p}$  disebut  $lwX_{(i)p}$  dan  $upX_{(i)p}$ .

Langkah 9. Hitung nilai dari  $a_{1p} = \text{abs}(X_{(1)p} - X_{(2)p})$ ,  $a_{2p} = \text{abs}(X_{(2)p} - X_{(3)p})$ ,  $c_p = \text{abs}(X_{(1)p} - X_{(3)p})$  dan  $d_p = \text{abs}(a_{1p} - X_{(3)p})$   $d_p = \text{abs}(a_{1p} - a_{2p})$  untuk masing-masing ID gen ke- $p$ .

Langkah 10. Identifikasi apakah nilai dari  $c_p < c_{\text{batas1}}$ . Jika benar, maka distribusi ID gen ke- $p$  adalah unimodal.

Langkah 11. Lakukan set data  $\mathbf{x}$  yang akan dianalisis sebagai  $c(x_1, x_2, \dots, x_n)$  dengan informasi ukuran sampel  $n$  sebagai  $\text{lenth}(\mathbf{x})$ .

Langkah 12. Menentukan inisialisasi sesuai dengan kondisi data untuk nilai  $\mathbf{w}$ ,  $\mu$ ,  $\sigma^2$  dan  $\mathbf{z}$  yang dibangun dengan fungsi *mixture* normal  $(w[i] * \text{dnorm}(x, \mu[i], \sqrt{\text{sigma2}[i]}))$  dengan iterasi nilai *sweep* sebanyak 200.000.

Langkah 13. Jalankan program RJMCMC dengan prinsip *bird-death*, *split-merge* untuk data  $\mathbf{x}$  berdasarkan inisialisasi  $\mathbf{z}$ ,  $\mathbf{w}$ , fungsi data  $\mathbf{x}$  dan  $\text{maksimum}(\mathbf{x})$ . Dalam menjalankan program RJMCMC, sebagai inputnya adalah data  $\mathbf{x}$ , jumlah iterasi (jumlah *sweep*=200.000),  $k_{\max} = n$ , mulai  $k=1$ , nilai  $\mathbf{z}$ ,  $\mathbf{w}$ ,  $\mu$ ,  $\sigma^2$  dan fungsi *mixture* normal.



Langkah 14. Simpan hasil jumlah komponen *mixture* dari densitas normal yang berupa bilangan bulat (diskrit) dengan operasi matematika *round* di setiap iterasi (*sweep*).

Langkah 15. Hitung jumlah komponen *mixture* dari densitas normal data  $X$

dengan rumus 
$$\text{round}\left(\frac{\text{jumlah komponen mixture setiap iterasi}}{\text{banyaknya iterasi}}\right)$$

dimana banyaknya *sweep* sebesar 200.000.

#### 4.5. Skenario Pembangkitan Data Simulasi

Dalam penelitian ini akan dilakukan pembangkitan data simulasi berdasarkan data *Chickpea* di Indonesia dengan konsep data *driven*. Dalam hal ini, data simulasi yang dibangkitkan memiliki ukuran sampel  $n = 3$ . Dalam kasus ini, data simulasi dibangkitkan mengikuti asumsi distribusi normal unimodal sesuai dengan informasi awal tentang data *Chickpea* yang diketahui melalui penelitian-penelitian sebelumnya dan untuk kemudahan pendekatan distribusi data. Untuk itu dibutuhkan sebuah skenario pembangkitan data simulasi yang dijabarkan dalam bentuk algoritma untuk mendapatkan data simulasi sesuai dengan pola data *Chickpea* kasus di Indonesia yang disebut sebagai Algoritma Pembangkitan Data Simulasi (Algoritma 4.5) yang dibangun oleh penulis (Astuti dkk., 2015c). Pembangkitan data simulasi ini dilakukan untuk melihat sejauh mana variabilitas data yang bisa dibangkitkan akan tetap mengikuti pola data *Chickpea*. Selain itu juga ingin mengidentifikasi apakah distribusi data *Chickpea* masih tetap mengikuti distribusi normal unimodal (sesuai informasi awal tentang data *Chickpea* yang diketahui) ataukah mengikuti distribusi *mixture* normal. Penetapan besarnya toleransi selisih antar data dalam setiap ID gen yang diamati sebesar 0,0999 yang ditentukan secara *trial* agar diperoleh nilai-nilai data simulasi yang tetap mengikuti pola dari data *Chickpea*.

Secara lengkap, skenario pembangkitan data simulasi dijabarkan dalam lima langkah Algoritma 4.5.

#### Algoritma 4.5. Algoritma Pembangkitan Data Simulasi

- Langkah 1. Pandang data  $\mathbf{x}$  sebagai variabel acak dengan  $x$  sebagai data pengamatan univariat dari data *Chickpea* dengan ukuran sampel  $n = 3$ , sehingga data pengamatannya adalah  $x_1, x_2, x_3$ .
- Langkah 2. Urutkan data pengamatan sehingga diperoleh  $x_{(1)}, x_{(2)}, x_{(3)}$ .
- Langkah 3. Hitung rata-rata ( $\bar{x}$ ) dan simpangan baku ( $s_x$ ) dari data  $\mathbf{x}$ .
- Langkah 4. Tetapkan toleransi selisih ( $\varepsilon$ ) sebesar 0,0999 (penetapan secara *trial*) agar data simulasi yang diperoleh tetap sepolanya dengan data aslinya. Penetapan nilai toleransi ini dilakukan dengan cara menghitung selisih absolut antara data *Chickpea* yang sudah terurut  $x_{(i)}$  dan data simulasi yang sudah terurut  $\hat{x}_{(i)}$ , yaitu  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon$ . Jika nilai selisih ( $\varepsilon$ ) ini terlalu besar atau terlalu kecil maka data simulasi akan melenceng dari data aslinya.
- Langkah 5. Lakukan pembangkitan data dengan cara sebagai berikut:
- Untuk perulangan  $i = 1$  atau  $i = 3$ , maka
    - (1). Bangkitkan  $\hat{x}_{(i)} \sim N(\bar{x}, s_x)$
    - (2). Hitung toleransi selisih  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon$
    - (3). Ulangi bangkitan pada (1) sampai tercapai kondisi  $\hat{x}_{(1)} \geq x_{(1)}$  dan  $\hat{x}_{(3)} \leq x_{(3)}$  dan  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon = 0,0999$  dapat terpenuhi.
  - Untuk perulangan  $i = 2$ , maka
    - (1). Bangkitkan  $\hat{x}_{(i)} \sim N(\bar{x}, s_x)$
    - (2). Hitung toleransi selisih  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon$
    - (3). Ulangi bangkitan pada (1) sampai tercapai kondisi  $|x_{(i)} - \hat{x}_{(i)}| < \varepsilon = 0,0999$  dapat terpenuhi.

#### 4.6. Goodness of Fit Kolmogorov-Smirnov (KS)

Untuk memvalidasi data simulasi yang telah dibangun apakah mengikuti data *Chickpea* digunakan *goodness of fit* Kolmogorov-Smirnov (KS). Penggunaan uji validasi *goodness of fit* KS memiliki beberapa asumsi yang harus dipenuhi yaitu sebuah sampel acak saling bebas  $\mathbf{x}$  dengan ukuran sampel sebanyak  $n$

dengan data pengamatan  $x_1, x_2, \dots, x_n$  dengan fungsi distribusi peluang kumulatif yang dihipotesiskan adalah  $F(x)$ . Untuk melakukan validasi pada data simulasi dengan *goodness of fit* KS melalui tahapan-tahapan berikut:

1. Tetapkan taraf nyata pengujian ( $\alpha = 5\%$ )
2. Tetapkan hipotesis pengujian:

$$H_0 : F(x) = F_0(x) \text{ untuk semua nilai } x \text{ (data berdistribusi } F_0(x))$$

dan

$$H_1 : F(x) \neq F_0(x) \text{ untuk sekurang-kurangnya satu nilai } x \text{ (data tidak berdistribusi } F_0(x))$$

3. Hitung statistik uji D

(1). Hitung distribusi empiris (fungsi distribusi sampel), yang dinyatakan dengan  $S(x)$  dimana  $S(x)$  adalah banyaknya nilai pengamatan dalam sampel yang kurang dari atau sama dengan  $x$  dibagi dengan jumlah sampel  $n$  (data diurutkan terlebih dahulu).

(2). Hitung distribusi dugaan (yang dihipotesiskan), untuk setiap  $i = 1, 2, \dots, n$ , yaitu:

$$F_0(x_i) = \int_{-\infty}^{x_i} f_0(x) dx.$$

Distribusi peluang untuk  $f_0(x)$  bisa unimodal maupun *mixture* sesuai distribusi data yang dihipotesiskan.

(3). Hitung nilai  $|S(x_i) - F_0(x_i)|$

(4). Hitung nilai  $|S(x_{i-1}) - F_0(x_i)|$ , untuk  $i = 1, 2, \dots, n$ , dimana nilai

$$|S(x_{i-1}) - F_0(x_i)| = F_0(x_i)$$

(5). Hitung nilai  $D = \max\left(\left(|S(x_i) - F_0(x_i)|\right), \left(|S(x_{i-1}) - F_0(x_i)|\right)\right)$

4. Pengambilan keputusan:

Gagal tolak hipotesis nol jika nilai  $D \leq D_{tabel}$  dimana  $D_{tabel}$  merupakan kuantil-kuantil statistik uji Kolmogorov-Smirnov dengan taraf nyata ( $\alpha$ ) dan jumlah pengamatan sebesar  $n$ , dimana nilai  $D_{tabel} = 0,708$  untuk  $\alpha = 0,05$  dan  $n=3$ .

Berdasarkan algoritma-algoritma yang telah dibangun dan uji validasi dari data simulasi, selanjutnya akan diimplementasikan pada data simulasi dan data *Chickpea*.

#### 4.7. Data Simulasi

Data simulasi dibangun dengan konsep data *driven* berdasarkan data *Chickpea* kasus di Indonesia. Pembangkitan data dilakukan pada setiap ID gen yang diamati dimana pada kasus ini adalah data LY-24 yang berjumlah 26 ID gen yang terdiri atas 15 ID gen dengan fungsi ketahanan dan 11 ID gen dengan fungsi energi. Pembangkitan data diulang sebanyak 1.000 kali dengan masing-masing perulangan di desain untuk ukuran sampel  $n=3$ . Sebagai contoh data ID gen LS0024, ukuran sampel  $n=3$ , maka akan diperoleh data bangkitan sebanyak 1.000 dengan masing-masing bangkitan data memiliki ukuran sampel  $n=3$ . Demikian seterusnya sampai dengan ID ke-26, dengan cara yang sama, proses pembangkitan data dilakukan untuk setiap ID gen yang diamati. Berikut ini disajikan deskripsi dari data *Chickpea* kasus di Indonesia.

##### a. Deskripsi Data *Chickpea*

Penelitian Harijati (2007) menggunakan data ekspresi gen dari jaringan tanaman *Chickpea* pada kondisi sehat (tanaman yang dibiarkan tumbuh dengan baik) dan pada kondisi sakit (tanaman yang diinokulasi penyakit ganas jamur patogen *Ascochyta Rabiei*). Serangan penyakit jamur patogen *Ascochyta Rabiei* ini sangat ganas, karena dapat menurunkan produksi tanaman *Chickpea* hingga 50%. Pada penelitian data *Chickpea* ini digunakan jenis tanaman *Chickpea Lasseter* (L) muda yang merupakan jenis tidak tahan (*susceptible*) terhadap serangan penyakit *Ascochyta Rabiei*. Tanaman *Chickpea* dipilih karena tanaman ini merupakan *legume* ketiga yang banyak dibutuhkan di dunia yaitu tanaman yang merupakan anggota dari famili *leguminosae (fabaceae)* yang buahnya berbentuk polong, bijinya kaya protein, akarnya mengandung bakteri simbiosis pengikat nitrogen dari udara.

Pengamatan data *Chickpea* dilakukan pengulangan sebanyak tiga kali dengan antar pengulangan dilakukan secara bebas (*independent*). Artinya antar

pengulangan menggunakan tanaman yang berbeda. Selain itu, pengamatan waktu serangan dilakukan pada tiga kelompok waktu yang berbeda, yaitu 24 jam setelah inokulasi (24 hpi-*hour post inoculation*), 48 jam setelah inokulasi (48 hpi) dan 72 jam setelah inokulasi (72 hpi). Efek serangan penyakit *Ascochyta Rabiei* mengakibatkan kondisi tanaman menjadi sakit (sebagai contoh jika terjadi pada jaringan daun, maka daun itu akan mulai berubah warna menjadi menguning). Hal ini dilakukan dengan tujuan agar dapat diketahui jenis ID gen dari tanaman *Chickpea* yang tahan maupun yang tidak tahan terhadap serangan jamur patogen *Ascochyta Rabiei*. Informasi ini sangat dibutuhkan agar dapat dilakukan penyisipan gen yang tahan terhadap serangan jamur patogen *Ascochyta Rabiei*. Hal ini tentu sangat bermanfaat dalam meningkatkan produksi tanaman *Chickpea* yang dibutuhkan di dunia. Dalam hal ini akan diidentifikasi kelompok-kelompok ID gen dalam *Up-regulated* (ID gen yang tahan dari serangan penyakit dan sangat signifikan dapat meningkatkan produksi tanaman *Chickpea*), *Regulated* (ID gen yang tidak cukup tahan dari serangan penyakit sehingga kurang mampu meningkatkan produksi tanaman *Chickpea*), dan *Down-regulated* (ID gen yang tidak tahan terhadap serangan penyakit dan sangat signifikan dapat menurunkan produksi tanaman *Chickpea*).

Hasil penelitian data *Chickpea* menunjukkan bahwa ada ribuan ID gen (sekitar 4.000 ID gen) yang dapat diamati pada jaringan tanaman *Chickpea*, tetapi yang mampu berekspresi dengan baik hanya sejumlah 62 ID gen. Berdasarkan 62 ID gen tersebut dapat dikelompokkan menjadi beberapa fungsi ID gen dalam jaringan tanaman *Chickpea*. Setiap ID gen yang diamati diulang sebanyak tiga kali secara independen, yaitu U1, U2, dan U3. Pengelompokan data perbedaan ekspresi gen pada jaringan tanaman *Chickpea* jenis *Lasseter* muda disimbolkan dengan LY-24 (U1, U2, U3).

Data perbedaan ekspresi gen yang disajikan secara lengkap pada Tabel 4.1 untuk ID gen dengan fungsi ketahanan dan Tabel 4.2 untuk ID gen dengan fungsi energi. Ada 15 ID gen yang memiliki fungsi ketahanan dan 11 ID gen dengan fungsi energi pada tanaman *Chickpea* pada kelompok *Lasseter* Muda (LY) pada kondisi sehat dan kondisi sakit. Kondisi sakit diamati setelah serangan selama 24 jam (LY-24). Pengamatan dilakukan secara berpasangan antara kondisi sakit dan

kondisi sehat dengan ukuran sampel sebanyak tiga (U1, U2, dan U3) pada setiap ID gen. Data yang disajikan adalah data perbedaan ekspresi gen yang merupakan ln rasio dari nilai ekspresi gen kondisi sakit terhadap nilai ekspresi gen kondisi

sehat yaitu:  $\ln \left( \frac{\text{nilai ekspresi gen sakit}}{\text{nilai ekspresi gen sehat}} \right)$ .

Hasil perhitungan data perbedaan ekspresi dari 15 ID gen dengan fungsi ketahanan dan 11 ID gen dengan fungsi energi menurut Tabel 4.1 dan Tabel 4.2, data tersebut selanjutnya dapat dikelompokkan menjadi tiga, yaitu bernilai negatif (-), nol, dan positif (+). Definisi setiap kelompok ID gen tersebut dapat dijabarkan sebagai berikut:

1. **Nilai negatif (-):** nilai ekspresi gen kondisi sehat lebih besar dari kondisi sakit. Hal ini berarti bahwa ID gen pada kelompok ini masuk dalam kelompok *Up-regulated*. Pada kondisi ini, ID gen tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan dapat meningkatkan secara signifikan dari produksi tanaman *Chickpea*. ID gen-ID gen yang masuk dalam kelompok ini dapat digunakan sebagai dasar untuk menemukan varietas baru tanaman *Chickpea* yang tahan penyakit melalui penyisipan gen. Sebagai contoh, jika serangan penyakit terjadi pada ID gen kelompok ini pada jaringan daun, maka daun tanaman *Chickpea* akan tetap segar dan sehat (biasanya ditunjukkan dengan warna daun yang hijau dan subur).
2. **Nilai positif (+):** nilai ekspresi gen kondisi sakit lebih besar dari kondisi sehat. Hal ini berarti bahwa ID gen pada kelompok ini masuk dalam kelompok *Down-regulated*. Pada kondisi ini, ID gen sangat tidak tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan dapat menurunkan secara signifikan dari produksi tanaman *Chickpea*. ID gen-ID gen yang masuk dalam kelompok ini tidak dapat digunakan sebagai dasar untuk menemukan varietas baru tanaman *Chickpea* yang tahan penyakit melalui penyisipan gen. Sebagai contoh, jika serangan penyakit terjadi pada ID gen kelompok ini pada jaringan daunnya, maka daun tanaman *Chickpea* akan menjadi layu dan sakit (biasanya ditunjukkan dengan warna daun yang menguning dan rontok).

3. **Nilai nol (0):** nilai ekspresi gen kondisi sakit sama atau relatif sama dengan kondisi sehat. Hal ini berarti bahwa ID gen pada kelompok ini masuk dalam kelompok *Regulated*. Pada kondisi ini, ID gen tersebut tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei*, sehingga tidak signifikan berpengaruh terhadap produksi tanaman *Chickpea*. ID gen-ID gen yang masuk dalam kelompok ini dapat digunakan sebagai media untuk menemukan varietas baru tanaman *Chickpea* yang tahan penyakit melalui penyisipan gen dari kelompok *Up-regulated*. Sebagai contoh, jika serangan penyakit terjadi pada ID gen kelompok ini pada jaringan daun tanaman *Chickpea*, maka serangan penyakit ini tidak akan menyebabkan adanya perubahan kondisi daun tanaman *Chickpea*.

Berdasarkan deskripsi data *Chickpea* kasus di Indonesia akan dilakukan pembangkitan data simulasi mengikuti pola data *Chickpea* berdasarkan konsep data *driven* menggunakan Algoritma 4.5 yang telah dijabarkan sebelumnya.

**Tabel 4.1.** Data Perbedaan Ekspresi Gen pada 15 ID Gen LY-24 dengan Fungsi Ketahanan

No	ID Gen	Dugaan Fungsi Gen dalam Jaringan Tanaman <i>Chickpea</i>	Ulangan			Rata-Rata
			U1	U2	U3	
1	LS0024	Dis.resist response prot. DRRG49-C	-0,6774	-0,4635	-0,0345	-0,3918
2	LS0035	Putative auxin-repressed prot.	-0,6801	-0,3246	-0,4733	-0,4926
3	LS0616	Putative auxin-repressed prot.	-0,9491	-0,3742	-0,3835	-0,5689
4	LS0159	Beta-glucan binding protein	-1,1699	-0,6513	-0,4797	-0,7670
5	LS0162	Pathogenesis related prot.	-0,5373	-0,7505	-0,0411	-0,4430
6	LS0185	Pathogenesis related prot.	-0,6855	-0,5024	-0,0483	-0,4121
7	LS0612	Laccase-like protein	-1,2309	-1,2487	-1,1705	-1,2167
8	LS0081	Pathogenesis related prot. 4A	-0,9387	-0,8640	-0,3513	-0,7180
9	LS0752	Pathogenesis related prot. 4A	-1,0805	-1,0393	-0,5859	-0,9019
10	LS0759	Pathogenesis related prot. 4A	-0,8109	-0,7626	-0,7468	-0,7734
11	LS0688	Pathogenesis related prot. 4A	-0,9052	-1,1297	-0,4464	-0,8271
12	U070	Homol to Avr9/Cf9 rapidly elic.prot.	-0,7078	-1,5376	-0,1971	-0,8142
13	U273	Prot.cont. leucine zipper motif	-1,9066	-1,2884	-1,3690	-1,5214
14	U278	Homol to SNAKIN2 antimicrobial peptide	-0,7286	-0,3576	-0,4814	-0,5225
15	U017	PR protein class 10	-0,3842	-0,2364	-0,1742	-0,2649

## b. Eksplorasi Data Simulasi

Eksplorasi data simulasi dilakukan terhadap data rata-rata hasil simulasi dari pembangkitan sebanyak 1.000 kali untuk mengetahui informasi lebih detail tentang data simulasi yang telah diperoleh, sehingga langkah-langkah analisis data selanjutnya dapat dipersiapkan. Penjabaran secara lengkap dari hasil eksplorasi data simulasi disajikan dalam Tabel 4.3 sampai dengan Tabel 4.6.

**Tabel 4.2.** Data Perbedaan Ekspresi Gen pada 11 ID Gen LY-24 dengan Fungsi Energi

No	ID Gen	Dugaan Fungsi Gen dalam Jaringan Tanaman <i>Chickpea</i>	Ulangan			Rata-Rata
			U1	U2	U3	
1	LS0132	Thioredoxin	-0,6167	-0,4241	-0,5351	-0,5253
2	LS0341	Thioredoxin H-type 1	-1,2333	-1,0169	-1,1079	-1,1193
3	U001	Photosystem I reaction center subunit IX	-0,9131	-0,5340	0,0680	-0,4597
4	U325	Photosystem I assembly prot.ycf3	-1,4996	-1,3232	-0,6091	-1,1440
5	U385	Chloroplast DNA	-0,9230	-0,4592	-0,2544	-0,5455
6	U447	ATP synthase C chain	-0,7630	-0,6241	-0,0346	-0,4739
7	CA0426	Chlorophyll a/b binding protein	-1,1077	-0,8226	8,08E-05	-0,6434
8	CA0534	Chlorophyll a/b binding protein	-0,8499	-0,8047	-0,5594	-0,7380
9	CA0702	NADH-plastoquinone oxidoreductase chain 1	-0,4451	-0,9759	-0,8224	-0,7478
10	CA0810	Chloroplast val-tRNA	-0,9457	-0,6581	-0,4086	-0,6708
11	CA1120	Chloroplast DNA	-0,7139	-0,7865	0,3875	-0,3710

### (1). Statistika Deskriptif Data Simulasi

Data simulasi yang telah dibangkitkan sesuai dengan karakteristik data *Chickpea*, artinya jika data *Chickpea* berdistribusi *mixture* normal dua atau tiga komponen ataupun berdistribusi normal unimodal, maka data simulasi dibangun sesuai dengan karakteristik data *Chickpea* tersebut. Sebagaimana yang telah dijelaskan dalam Bab III, bahwa pendekatan distribusi normal yang digunakan dalam penelitian ini mengikuti distribusi normal sebagaimana yang digunakan dalam penelitian-penelitian sebelumnya dan bukan dari hasil statistik uji kenormalan. Hal ini disebabkan oleh ukuran sampel data *Chickpea* sebagai acuan pembangkitan data simulasi memiliki jumlah yang sangat sedikit, yaitu  $n = 3$ , sehingga menghasilkan informasi kenormalan yang semu. Informasi ini dapat dijelaskan dari hasil uji kenormalan KS pada Tabel 4.11 dan Tabel 4.12, yaitu



distribusi setiap ID yang diamati dinyatakan berdistribusi normal tetapi memiliki nilai *skewness* dan nilai *kurtosis* bukan 0. Hasil identifikasi jumlah komponen *mixture* dari data *Chickpea* disajikan dalam Tabel 4.13. Secara lengkap, data simulasi dideskriptifkan dalam Tabel 4.3 sampai dengan Tabel 4.6.

Pada data simulasi, setiap ID gen dibangkitkan dengan ukuran sampel sebanyak tiga, yaitu U1, U2 dan U3 dan pembangkitan ini diulang sebanyak 1.000 kali. Sebagai contoh, pada Tabel 4.3, nilai U1=-0,6241 pada ID gen LS0024 berasal dari nilai rata-rata 1.000 kali ulangan. Sedangkan nilai rata-rata yang sebesar -0,3918 pada ID gen LS0024 berasal dari rata-rata tiga kali ulangan. Penyusunan data dalam Tabel 4.3 sampai dengan Tabel 4.6 dilakukan sesuai dengan fungsi ID gen dalam data *Chickpea* dan antara ID gen yang satu dengan ID gen yang lain bersifat independen. Deskripsi data simulasi pada Tabel 4.5 dan Tabel 4.6 digunakan untuk memberikan informasi secara umum dari data simulasi melalui statistika deskriptif, tetapi data simulasi yang nantinya di analisis tetap data perbedaan ekspresi gen setiap ID gen yang diperoleh pada rata-rata 1.000 kali pembangkitan di setiap ulangan U1, U2 dan U3.

## **(2). Plot Vertikal Data Simulasi**

Plot data simulasi secara vertikal dilakukan untuk mengetahui adanya indikasi *mixture* dalam data secara visual. Hasil plot disajikan pada Gambar 4.2 untuk ID gen fungsi ketahanan dan Gambar 4.3 untuk ID gen fungsi energi. Berdasarkan hasil plot vertikal data simulasi dapat diketahui bahwa dari 1.000 kali pembangkitan untuk setiap ulangan U1, U2 dan U3 pada setiap ID gen, memiliki hasil sesuai dengan pola data *Chickpea* dan secara visual ada indikasi *mixture* dalam data simulasi. Kesamaan pola data simulasi dengan pola data *Chickpea* ini telah dilakukan pengujian dengan *goodness of fit* KS sebagaimana ditunjukkan hasilnya pada Tabel 4.7 dan Tabel 4.8.

## **(3). Validasi Data Simulasi**

Untuk memvalidasi data simulasi yang telah dibangun apakah sesuai/mengikuti data *Chickpea* digunakan *goodness of fit* KS dua sample independen sebagaimana yang telah dijabarkan sebelumnya. Hasil yang diperoleh disajikan pada Tabel 4.7 dan Tabel 4.8.

Hasil validasi yang disajikan pada Tabel 4.7 dan Tabel 4.8 menunjukkan bahwa semua data simulasi yang dibangkitkan tetap sesuai dengan data *Chickpea*. Artinya, jika data *Chickpea* berdistribusi normal *mixture* dua komponen, maka data simulasi juga berdistribusi normal *mixture* dua komponen. Apabila data *Chickpea* berdistribusi normal *mixture* tiga komponen, maka data simulasi juga berdistribusi normal *mixture* tiga komponen. Demikian juga jika data *Chickpea* berdistribusi normal unimodal maka data simulasi juga berdistribusi normal unimodal. Kesesuaian hasil ini ditunjukkan oleh semua nilai  $P_{value} > 0,05$  untuk setiap ID gen yang diamati. Menurut Gambar 4.5 dan Gambar 4.6 dapat ditunjukkan bahwa nilai-nilai KS dari 1.000 kali pembangkitan data simulasi untuk setiap ID gen pada fungsi ketahanan dan fungsi energi tetap memiliki nilai  $P_{value} > 0,05$ . Model yang dibangun di setiap ID gen pada data simulasi disesuaikan dengan model yang paling tepat untuk data *Chickpea*. Hasil validasi data simulasi ini telah membuktikan bahwa ada beberapa data *Chickpea* yang masih tetap mengikuti distribusi normal unimodal (sesuai informasi awal tentang data *Chickpea* yang diketahui), dimana dalam hal ini ada tujuh data ID gen, yaitu: LS0035, LS0612, LS0759, U278, U017, LS0132 dan LS0341. Sedangkan 16 ID gen memiliki distribusi *mixture* normal bimodal (*mixture* normal dengan dua komponen), yaitu ID gen: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, U325, U385, U447, CA0426, CA0534, CA0702 dan CA1120 dan ada tiga ID gen yang memiliki distribusi *mixture* normal trimodal (*mixture* normal dengan tiga komponen), yaitu ID gen: U070, U001 dan CA0810. Selain itu, melalui data simulasi ini juga telah ditemukan nilai toleransi  $\varepsilon = 0,0999$  yang menunjukkan nilai variabilitas data simulasi yang masih mungkin dilakukan agar pola data simulasi tetap mengikuti pola data *Chickpea*.

#### **4.8. Data *Chickpea***

Data *Chickpea* merupakan data perbedaan ekspresi gen dari jaringan tanaman *Chickpea* sebagai hasil dari percobaan *microarray* yang merupakan hasil dari penelitian Harijati (2007). Eksplorasi data *Chickpea* secara lengkap disajikan dalam Tabel 4.9 dan Tabel 4.10 serta Gambar 4.3 dan Gambar 4.4.

#### a. Statistika Deskriptif Data *Chickpea*

Hasil statistika deskriptif data *Chickpea* untuk 15 ID gen dengan fungsi ketahanan dan 11 ID gen dengan fungsi energi disajikan dalam Tabel 4.9 dan Tabel 4.10. Statistika deskriptif yang disajikan terdiri dari ukuran sampel  $n$ , nilai rata-rata, nilai simpangan baku, nilai minimum, nilai Quartil1 (Q1), nilai Quartil3 (Q3), nilai maksimum, nilai kisaran (*range*) dan nilai *skewness* dari data. Masing-masing nilai statistika deskriptif diberikan untuk setiap ID gen yang diamati.

**Tabel 4.3.** Rata-Rata Data Simulasi dari Pembangkitan 1.000 kali pada setiap Ulangan untuk ID Gen LY-24 Fungsi Ketahanan

No	ID Gen Fungsi Ketahanan	Ulangan			Rata-Rata
		U1	U2	U3	
1	LS0024	-0,6241	-0,4632	-0,0880	-0,3918
2	LS0035	-0,6251	-0,4767	-0,3778	-0,4932
3	LS0616	-0,8970	-0,3885	-0,4253	-0,5703
4	LS0159	-1,1179	-0,6544	-0,5298	-0,7674
5	LS0162	-0,6988	-0,5336	-0,0943	-0,4422
6	LS0185	-0,6350	-0,4998	-0,1008	-0,4119
7	LS0612	-1,2127	-1,2197	-1,2122	-1,2127
8	LS0081	-0,8875	-0,8623	-0,4046	-0,7181
9	LS0752	-1,0311	-1,0316	-0,6381	-0,6876
10	LS0759	-0,7772	-0,7728	-0,7785	-0,7762
11	LS0688	-1,0795	-0,9007	-0,4983	-0,3599
12	U070	-1,4864	-0,7061	-0,2483	-0,4955
13	U273	-1,8522	-1,3727	-1,3413	-1,5221
14	U278	-0,6743	-0,4831	-0,4098	-0,5224
15	U017	-0,3296	-0,2470	-0,2276	-0,2681

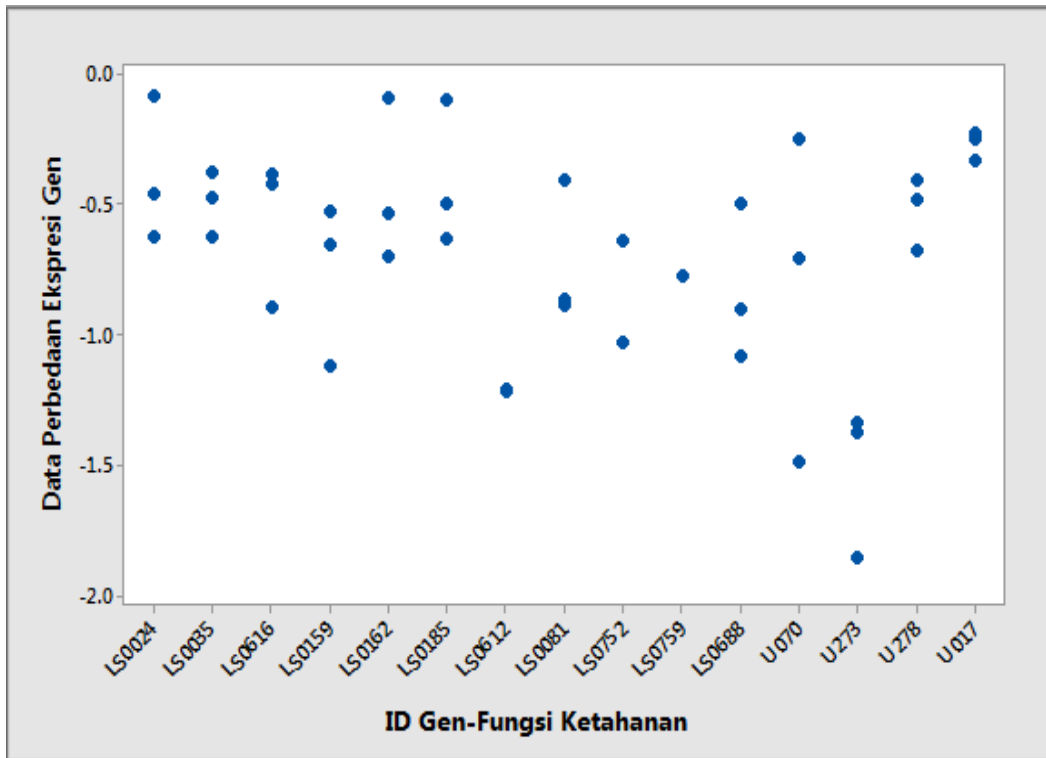
#### b. Distribusi Data *Chickpea*

Identifikasi bentuk distribusi data *Chickpea* untuk setiap ID gen dengan statistik uji kenormalan KS dilakukan untuk membuktikan bahwa statistik uji ini menghasilkan informasi kenormalan yang semu akibat dari ukuran sampel  $n = 3$  yang cukup kecil. Hasil identifikasi distribusi kenormalan dapat dilihat pada Tabel 4.11 dan Tabel 4.12. Berdasarkan hasil identifikasi distribusi kenormalan data *Chickpea* yang telah dilakukan tersebut dapat diketahui bahwa 26 ID gen yang

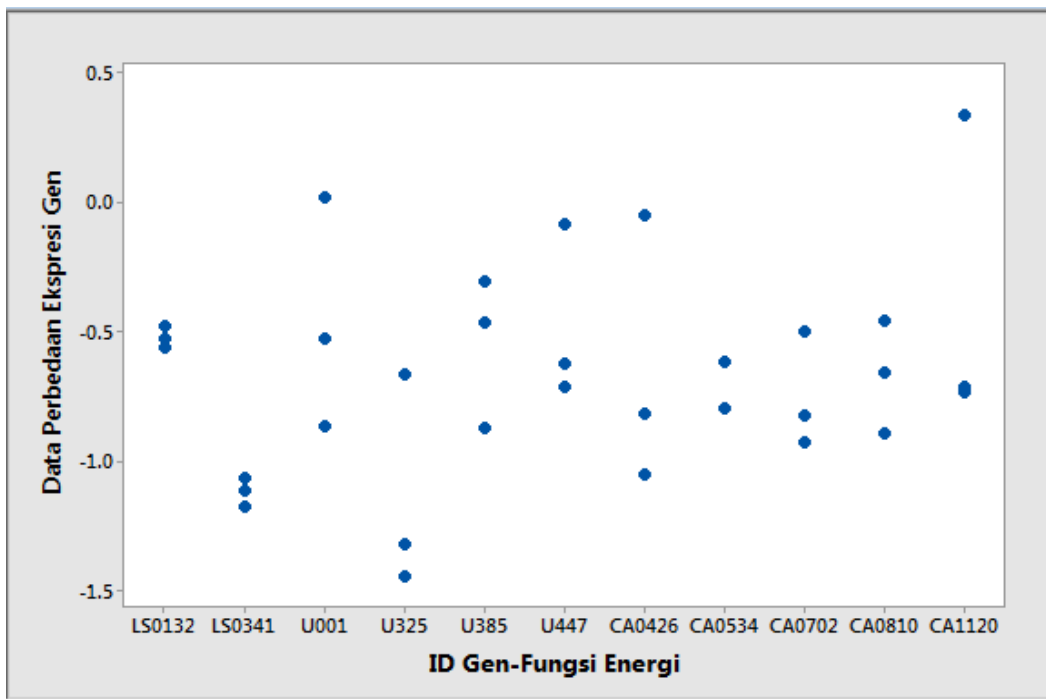
diamati memiliki distribusi normal. Hal ini dapat diketahui berdasarkan  $p_{value}$  yang diperoleh dari uji kenormalan Kolmogorov-Smirnov (KS) lebih besar atau sama dengan  $\alpha = 0,05$ . Hasil uji kenormalan KS ini memberikan informasi distribusi kenormalan yang semu karena ukuran sampel yang cukup kecil, yaitu  $n = 3$ . Hal ini dapat dilihat bahwa walaupun hasil uji distribusi kenormalan KS menunjukkan berdistribusi normal tetapi dari nilai *skewness* dan *kurtosis* yang dihasilkan memiliki nilai bukan 0. Artinya informasi distribusi normal yang diperoleh dari KS adalah informasi semu. Oleh karena itu, dalam penelitian ini pendekatan distribusi normal untuk data *Chickpea* bukan berdasarkan uji kenormalan KS tetapi mengikuti distribusi normal berdasarkan penelitian-penelitian data *microarray* sebelumnya.

**Tabel 4.4.** Rata-Rata Data Simulasi dari Pembangkitan 1.000 kali pada setiap Ulangan untuk ID Gen LY-24 Fungsi Energi

No	ID Gen Fungsi Energi	Ulangan			Rata-Rata
		U1	U2	U3	
1	LS0132	-0,5616	-0,5307	-0,4781	-0,5235
2	LS0341	-1,1792	-1,1134	-1,0687	-1,1204
3	U001	-0,8628	-0,5305	0,0158	-0,4592
4	U325	-1,4493	-1,3215	-0,6623	-0,9236
5	U385	-0,8709	-0,4621	-0,3058	-0,5463
6	U447	-0,7124	-0,6218	-0,0870	-0,4737
7	CA0426	-1,0557	-0,8208	-0,0495	-0,3519
8	CA0534	-0,7974	-0,7974	-0,6152	-0,7367
9	CA0702	-0,9262	-0,8219	-0,4988	-0,7490
10	CA0810	-0,8932	-0,6565	-0,4592	-0,6696
11	CA1120	-0,7348	-0,7129	0,3356	-0,3707



**Gambar 4.2.** Plot Vertikal Rata-Rata dari 1.000 Kali Pembangkitan Data Simulasi pada U1, U2 dan U3 untuk Setiap ID Gen Fungsi Ketahanan



**Gambar 4.3.** Plot Vertikal Rata-Rata dari 1.000 Kali Pembangkitan Data Simulasi pada U1, U2 dan U3 untuk Setiap ID Gen Fungsi Energi

**Tabel 4.5.** Statistika Deskriptif Data Simulasi (Pembangkitan 1.000 kali) untuk ID Gen LY-24 Fungsi Ketahanan

No	ID Gen Fungsi Ketahanan	$n$	Rata-Rata	Simpangan Baku	Minimum	Q1	Q3	Maksimum	Kisaran ( <i>Range</i> )	<i>Skewness</i>
1	LS0024	3	-0,3918	0,2773	-0,6241	-0,5436	-0,2756	-0,0880	0,5361	1,09
2	LS0035	3	-0,4932	0,1292	-0,6251	-0,5515	-0,4273	-0,3767	0,2484	-0,59
3	LS0616	3	-0,5703	0,2852	-0,8970	-0,6666	-0,4069	-0,3775	0,5195	-1,70
4	LS0159	3	-0,7674	0,3118	-1,1179	-0,8862	-0,5921	-0,5296	0,5882	-1,42
5	LS0162	3	-0,4422	0,3143	-0,6988	-0,6162	-0,3140	-0,0943	0,6045	1,20
6	LS0185	3	-0,4119	0,2798	-0,6351	-0,5674	-0,3003	-0,1008	0,5343	1,28
7	LS0612	3	-1,2149	0,0256	-1,2397	-1,2269	-1,2025	-1,1908	0,0489	-1,70
8	LS0081	3	-0,7181	0,2737	-0,9028	-0,8749	-0,6258	-0,4047	0,4982	1,72
9	LS0752	3	-0,9003	0,2292	-1,0580	-1,0314	-0,8214	-0,6382	0,4199	1,73
10	LS0759	3	-0,7762	0,0212	-0,7960	-0,7865	-0,7663	-0,7556	0,0405	1,37
11	LS0688	3	-0,8262	0,2996	-1,0795	-0,9901	-0,6995	-0,4983	0,5812	1,06
12	U070	3	-0,8136	0,6270	-1,4864	-1,0962	-0,4772	-0,2483	1,2381	-0,75
13	U273	3	-1,5221	0,2881	-1,8522	-1,6189	-1,3570	-1,3284	0,5238	-1,71
14	U278	3	-0,5224	0,1409	-0,6743	-0,5805	-0,4465	-0,4062	0,2681	-1,19
15	U017	3	-0,2681	0,0629	-0,3312	-0,2967	-0,2365	-0,2109	0,1203	-1,49

**Tabel 4.6.** Statistika Deskriptif Data Simulasi (Pembangkitan 1.000 kali) untuk ID Gen LY-24 Fungsi Energi

No	ID Gen Fungsi Energi	<i>n</i>	Rata-Rata	Simpangan Baku	Minimum	Q1	Q3	Maksimum	Kisaran ( <i>Range</i> )	<i>Skewness</i>
1	LS0132	3	-0,5235	0,0528	-0,5735	-0,5488	-0,4985	-0,4728	0,1007	0,75
2	LS0341	3	-1,1204	0,0654	-1,1828	-1,1513	-1,0892	-1,0585	0,1243	-0,56
3	U001	3	-0,4592	0,4450	-0,8628	-0,6966	-0,2574	0,0158	0,8786	0,70
4	U325	3	-1,1443	0,4237	-1,4493	-1,3854	-0,9919	-0,6623	0,7871	1,56
5	U385	3	-0,5462	0,2938	-0,8709	-0,6665	-0,3839	-0,3058	0,5651	-1,19
6	U447	3	-0,4737	0,3397	-0,7140	-0,6671	-0,3536	-0,0870	0,6269	1,59
7	CA0426	3	-0,6420	0,5275	-1,0557	-0,9382	-0,4351	-0,0495	1,0062	1,35
8	CA0534	3	-0,7367	0,1099	-0,8243	-0,7974	-0,7706	-0,6152	0,2091	1,73
9	CA0702	3	-0,7490	0,2256	-0,9271	-0,8740	-0,6599	-0,4985	0,4283	1,32
10	CA0810	3	-0,6696	0,2202	-0,8932	-0,7748	-0,5579	-0,4592	0,4340	-0,27
11	CA1120	3	-0,3707	0,6126	-0,7523	-0,7239	-0,1799	0,3356	1,0879	1,73

**Tabel 4.7.** Hasil Uji KS Dua Sampel Independen antara Data *Chickpea* dengan Data Simulasi untuk ID Gen Fungsi Ketahanan

No	ID Gen Fungsi Ketahanan	<i>n</i>	Model yang Dibangun sesuai Pola Data <i>Chickpea</i>	P-value
1	LS0024	3	<i>Mixture</i> Normal 2 Komponen	0,996
2	LS0035	3	Normal Unimodal	0,996
3	LS0616	3	<i>Mixture</i> Normal 2 Komponen	0,518
4	LS0159	3	<i>Mixture</i> Normal 2 Komponen	0,996
5	LS0162	3	<i>Mixture</i> Normal 2 Komponen	0,996
6	LS0185	3	<i>Mixture</i> Normal 2 Komponen	0,996
7	LS0612	3	Normal Unimodal	0,518
8	LS0081	3	<i>Mixture</i> Normal 2 Komponen	0,996
9	LS0752	3	<i>Mixture</i> Normal 2 Komponen	0,518
10	LS0759	3	Normal Unimodal	0,518
11	LS0688	3	<i>Mixture</i> Normal 2 Komponen	0,996
12	U070	3	<i>Mixture</i> Normal 3 Komponen	0,996
13	U273	3	<i>Mixture</i> Normal 2 Komponen	0,996
14	U278	3	Normal Unimodal	0,996
15	U017	3	Normal Unimodal	0,996

**Tabel 4.8.** Hasil Uji KS Dua Sampel Independen antara Data *Chickpea* dengan Data Simulasi untuk ID Gen Fungsi Energi

No	ID Gen Fungsi Energi	<i>n</i>	Model yang Dibangun sesuai Pola Data <i>Chickpea</i>	P-value
1	LS0132	3	Normal Unimodal	0,996
2	LS0341	3	Normal Unimodal	0,996
3	U001	3	<i>Mixture</i> Normal 3 Komponen	0,996
4	U325	3	<i>Mixture</i> Normal 2 Komponen	0,996
5	U385	3	<i>Mixture</i> Normal 2 Komponen	0,996
6	U447	3	<i>Mixture</i> Normal 2 Komponen	0,996
7	CA0426	3	<i>Mixture</i> Normal 2 Komponen	0,996
8	CA0534	3	<i>Mixture</i> Normal 2 Komponen	0,518
9	CA0702	3	<i>Mixture</i> Normal 2 Komponen	0,996
10	CA0810	3	<i>Mixture</i> Normal 3 Komponen	0,996
11	CA1120	3	<i>Mixture</i> Normal 2 Komponen	0,996



### c. Plot Vertikal Data *Chickpea*

Plot secara vertikal dari data *Chickpea* digunakan untuk kemudahan identifikasi indikasi distribusi *mixture* dalam data. Hasil plot secara vertikal disajikan dalam Gambar 4.3 untuk ID gen fungsi ketahanan dan Gambar 4.4 untuk ID gen fungsi energi.

Berdasarkan plot data secara vertikal sebagaimana ditunjukkan pada Gambar 4.3 dan Gambar 4.4 serta penetapan pendekatan distribusi normal yang digunakan dapat diketahui bahwa ada indikasi data berdistribusi normal unimodal, *mixture* normal bimodal, dan *mixture* normal trimodal. Ada tujuh ID gen yang memiliki indikasi berdistribusi normal unimodal, yaitu ID gen: LS0035, LS0612, LS0759, U278, U017, LS0132 dan LS0341. Ada 16 ID gen yang memiliki indikasi berdistribusi *mixture* normal bimodal (*mixture* normal dengan dua komponen), yaitu ID gen: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, U325, U385, U447, CA0426, CA0534, CA0702 dan CA1120. Kemudian ada tiga ID gen yang memiliki indikasi berdistribusi *mixture* normal trimodal (*mixture* normal dengan tiga komponen), yaitu ID gen: U070, U001 dan CA0810.

### d. Deteksi Jumlah Komponen *Mixture* Data *Chickpea*

Deteksi jumlah komponen *mixture* dalam data *Chickpea* menggunakan Algoritma 4.3, yaitu algoritma tentang Pengembangan Algoritma RJMCMC. Secara lengkap, hasil identifikasi jumlah komponen *mixture* dalam data *Chickpea* disajikan pada Tabel 4.13.

Berdasarkan Tabel 4.11, Tabel 4.12, dan Tabel 4.13 dapat ditunjukkan bahwa dari 26 ID gen yang diamati (15 ID gen dengan fungsi ketahanan dan 11 ID gen dengan fungsi energi) teridentifikasi ada lima ID gen dari fungsi ketahanan berdistribusi normal unimodal, yaitu ID gen: LS0035, LS0612, LS0759, U278 dan U017, ada sembilan ID gen berdistribusi *mixture* normal dengan dua komponen, yaitu ID gen: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688 dan U273 serta ada satu ID gen berdistribusi *mixture* normal dengan tiga komponen, yaitu ID gen U070. Sedangkan untuk ID gen dengan fungsi energi, teridentifikasi ada dua ID gen berdistribusi normal unimodal, yaitu ID gen:

LS0132 dan LS0341, ada tujuh ID gen berdistribusi *mixture* normal dengan dua komponen, yaitu ID gen: U325, U385, U447, CA0426, CA0534, CA0702 dan CA1120, serta ada dua ID gen berdistribusi *mixture* normal dengan tiga komponen, yaitu ID gen U001 dan CA0810.

Secara umum, dari 26 ID gen yang diamati, ada tujuh ID gen yang memiliki distribusi normal unimodal, ada 16 ID gen yang memiliki distribusi *mixture* normal bimodal (*mixture* normal dengan dua komponen) dan ada tiga ID gen yang memiliki distribusi *mixture* normal trimodal (*mixture* normal dengan tiga komponen). Hasil uji Tabel 4.13 ini digunakan sebagai dasar pembangkitan data simulasi untuk setiap ID gen yang diamati.

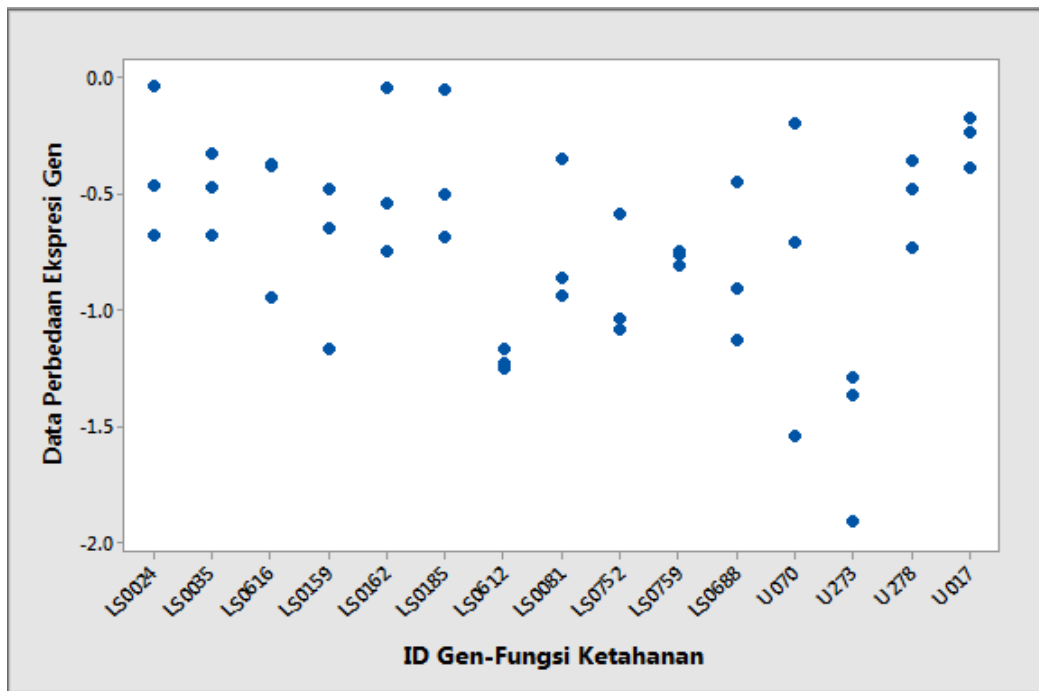
Hasil eksplorasi, identifikasi distribusi dan identifikasi jumlah komponen *mixture* dalam data yang diamati, baik untuk data simulasi dan data *Chickpea* yang telah dilakukan dan selanjutnya akan dibangun pemodelan normal dengan pendekatan Bayesian Unimodal Model (BUM), Bayesian Model *Averaging* (BMA), Bayesian *Mixture* Model (BMM), dan Bayesian *Mixture* Model *Averaging* (BMMA), dimana masing-masing model akan dibandingkan dengan pemodelan OLS berdasarkan nilai *goodness of fit* KS. Pemodelan BUM dan BMA akan disajikan dalam BAB V, Pemodelan BMM akan disajikan dalam BAB VI, dan Pemodelan BMMA akan disajikan dalam BAB VII.

**Tabel 4.9.** Statistika Deskriptif Data *Chickpea*, Perbedaan Ekspresi Gen dari 15 ID Gen LY-24 dengan Fungsi Ketahanan

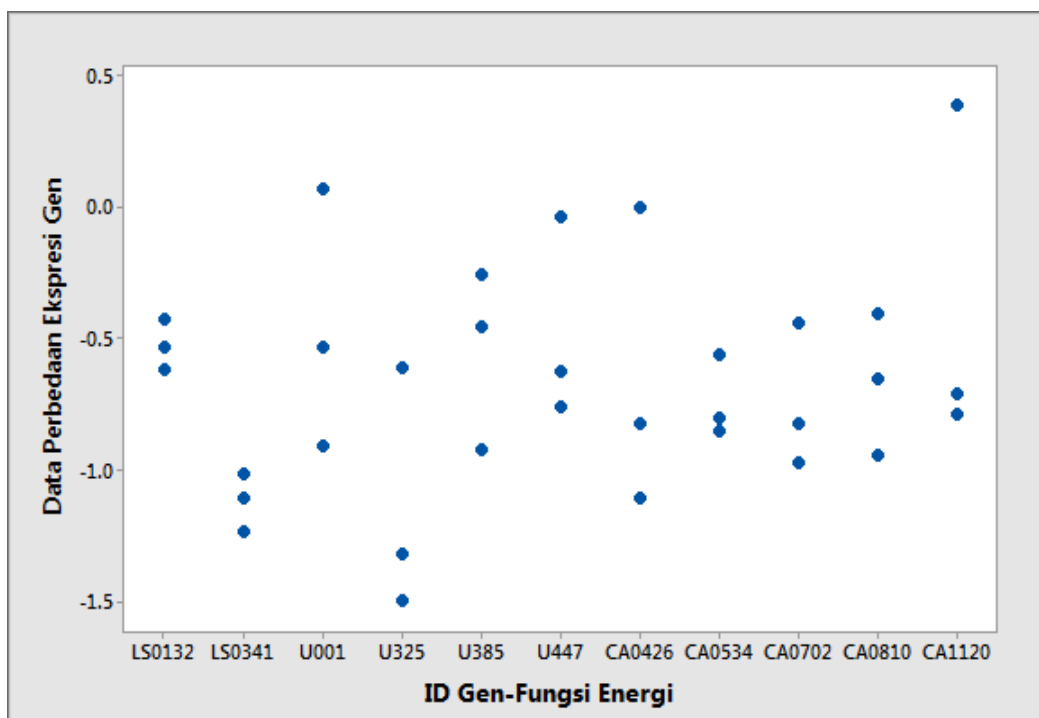
No	ID Gen Fungsi Ketahanan	<i>n</i>	Rata-Rata	Simpangan Baku	Minimum	Q1	Q3	Maksimum	Kisaran ( <i>Range</i> )	<i>Skewness</i>
1	LS0024	3	-0.392	0.327	-0.677	-0.677	-0.034	-0.034	0.643	0,94
2	LS0035	3	-0.493	0.179	-0.680	-0.680	-0.325	-0.325	0.355	-0,48
3	LS0616	3	-0.569	0.329	-0.949	-0.949	-0.374	-0.374	0.575	-1,73
4	LS0159	3	-0.767	0.359	-1.170	-1.170	-0.480	-0.480	0.690	-1,30
5	LS0162	3	-0.443	0.364	-0.750	-0.750	-0.041	-0.041	0.709	1,09
6	LS0185	3	-0.412	0.328	-0.686	-0.686	-0.048	-0.048	0.637	1,15
7	LS0612	3	-1.217	0.041	-1.249	-1.249	-1.171	-1.171	0.078	1,37
8	LS0081	3	-0.718	0.320	-0.939	-0.939	-0.351	-0.351	0.587	1,63
9	LS0752	3	-0.902	0.274	-1.081	-1.081	-0.586	-0.586	0.495	1,69
10	LS0759	3	-0.773	0.033	-0.811	-0.811	-0.747	-0.747	0.064	-1,31
11	LS0688	3	-0.827	0.348	-1.130	-1.130	-0.446	-0.446	0.683	0,96
12	U070	3	-0.814	0.677	-1.538	-1.538	-0.197	-0.197	1.340	-0,69
13	U273	3	-1.521	0.336	-1.907	-1.907	-1.288	-1.288	0.618	-1,62
14	U278	3	-0.523	0.189	-0.729	-0.729	-0.358	-0.358	0.371	-0,93
15	U017	3	-0.265	0.108	-0.384	-0.384	-0.174	-0.174	0.210	-1,11

**Tabel 4.10.** Statistika Deskriptif Data *Chickpea*, Perbedaan Ekspresi Gen dari 11 ID Gen LY-24 dengan Fungsi Energi

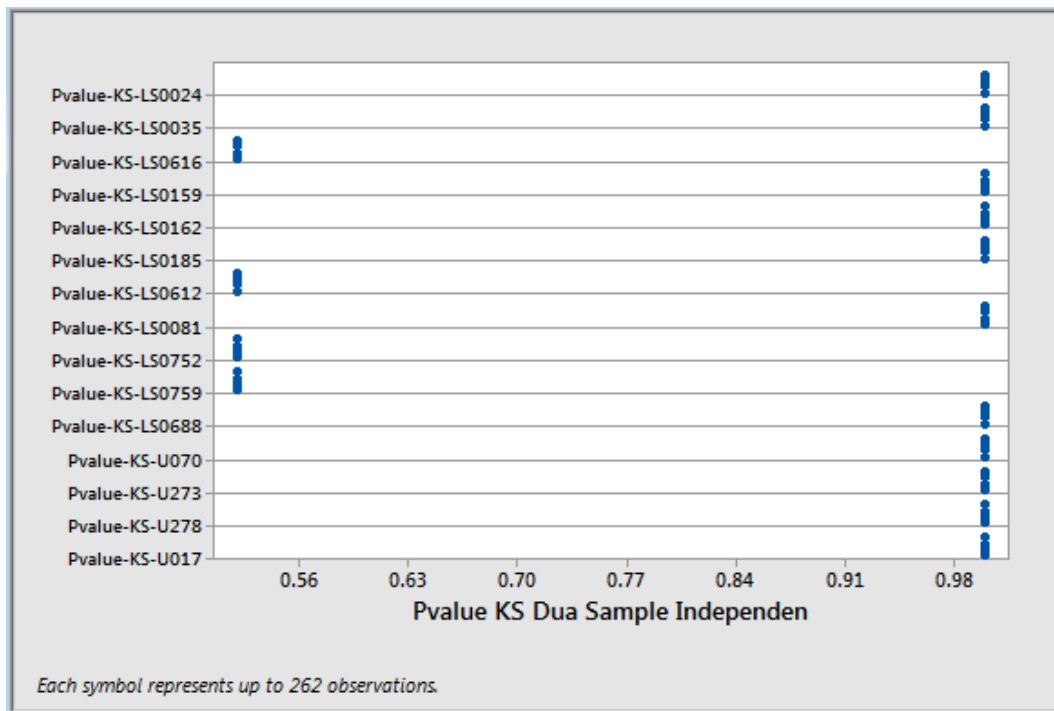
No	ID Gen Fungsi Energi	<i>n</i>	Rata-Rata	Simpangan Baku	Minimum	Q1	Q3	Maksimum	Kisaran ( <i>Range</i> )	<i>Skewness</i>
1	LS0132	3	-0.525	0.097	-0.617	-0.617	-0.424	-0.424	0.193	0,45
2	LS0341	3	-1.119	0.109	-1.233	-1.233	-1.017	-1.017	0.216	-0,47
3	U001	3	-0.460	0.495	-0.913	-0.913	0.068	0.068	0.981	0,66
4	U325	3	-1.144	0.472	-1.500	-1.500	-0.609	-0.609	0.890	1,46
5	U385	3	-0.546	0.343	-0.923	-0.923	-0.254	-0.254	0.669	-1,06
6	U447	3	-0.474	0.387	-0.763	-0.763	-0.035	-0.035	0.728	1,48
7	CA0426	3	-0.643	0.575	-1.108	-1.108	0.000	0.00008	1.108	1,27
8	CA0534	3	-0.738	0.156	-0.850	-0.850	-0.559	-0.559	0.291	1,57
9	CA0702	3	-0.748	0.273	-0.976	-0.976	-0.445	-0.445	0.531	1,14
10	CA0810	3	-0.671	0.269	-0.946	-0.946	-0.409	-0.409	0.537	-0,21
11	CA1120	3	-0.371	0.658	-0.786	-0.786	0.387	0.387	1.174	1,71



**Gambar 4.4.** Plot Vertikal Data *Chickpea* untuk 15 ID Gen Fungsi Ketahanan

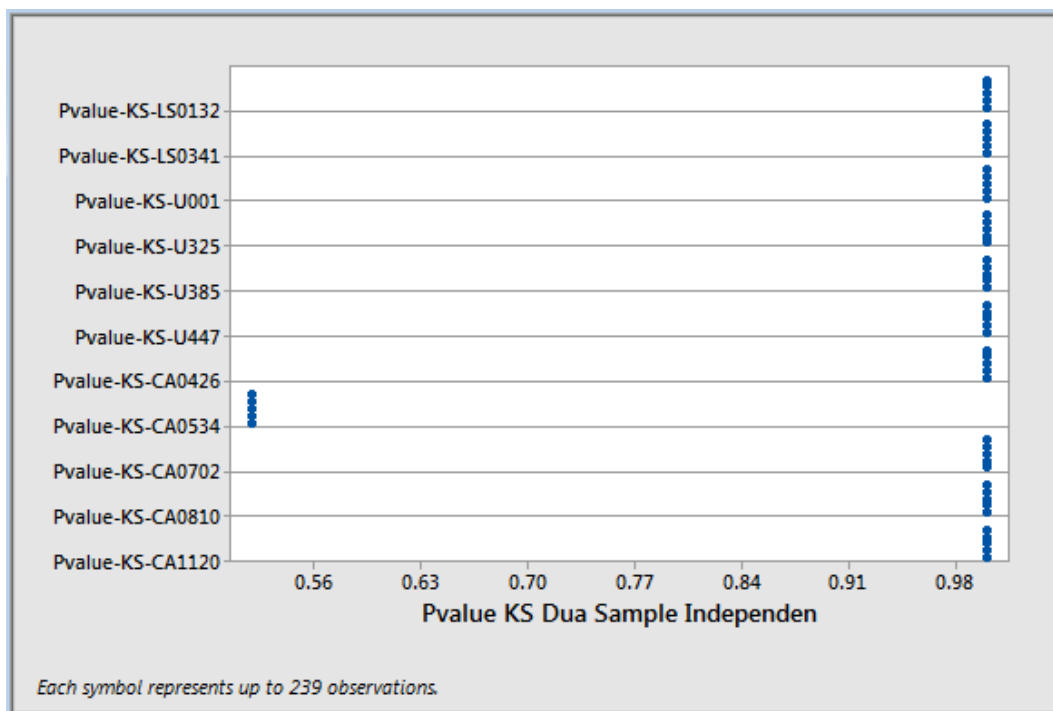


**Gambar 4.5.** Plot Vertikal Data *Chickpea* untuk 11 ID Gen Fungsi Energi



$\alpha = 0,05$

**Gambar 4.6.** Dotplot Nilai P-value KS dari 1.000 Kali Pembangkitan Data Simulasi untuk Setiap ID Gen Fungsi Ketahanan



$\alpha = 0,05$

**Gambar 4.7.** Dotplot Nilai P-value KS dari 1.000 Kali Pembangkitan Data Simulasi untuk Setiap ID Gen Fungsi Energi

**Tabel 4.11.** Uji Distribusi Normal Data *Chickpea* untuk 15 ID Gen Fungsi Ketahanan

No	ID Gen	Ulangan			Rata-Rata	Skewness	Kurtosis	$P_{value}$ Statistik Uji KS
		U1	U2	U3				
1	LS0024	-0,6774	-0,4635	-0,0345	-0,3918	0,94	*	> 0,150
2	LS0035	-0,6801	-0,3246	-0,4733	-0,4926	-0,48	*	> 0,150
3	LS0616	-0,9491	-0,3742	-0,3835	-0,5689	-1,73	*	0,086
4	LS0159	-1,1699	-0,6513	-0,4797	-0,7670	-1,30	*	> 0,150
5	LS0162	-0,5373	-0,7505	-0,0411	-0,4430	1,09	*	> 0,150
6	LS0185	-0,6855	-0,5024	-0,0483	-0,4121	1,15	*	> 0,150
7	LS0612	-1,2309	-1,2487	-1,1705	-1,2167	1,37	*	> 0,150
8	LS0081	-0,9387	-0,8640	-0,3513	-0,7180	1,63	*	> 0,150
9	LS0752	-1,0805	-1,0393	-0,5859	-0,9019	1,69	*	0,130
10	LS0759	-0,8109	-0,7626	-0,7468	-0,7734	-1,31	*	> 0,150
11	LS0688	-0,9052	-1,1297	-0,4464	-0,8271	0,96	*	> 0,150
12	U070	-0,7078	-1,5376	-0,1971	-0,8142	-0,69	*	> 0,150
13	U273	-1,9066	-1,2884	-1,3690	-1,5214	-1,62	*	> 0,150
14	U278	-0,7286	-0,3576	-0,4814	-0,5225	-0,93	*	> 0,150
15	U017	-0,3842	-0,2364	-0,1742	-0,2649	-1,11	*	> 0,150

\* Nilai kurtosis tidak teridentifikasi,  $\alpha = 0,05$

**Tabel 4.12.** Uji Distribusi Normal Data *Chickpea* untuk 11 ID Gen Fungsi Energi

No	ID Gen	Ulangan			Rata-Rata	Skewness	Kurtosis	$P_{value}$ Statistik Uji KS
		U1	U2	U3				
1	LS0132	-0,5616	-0,5307	-0,4781	-0,5235	0,45	*	> 0,150
2	LS0341	-1,1792	-1,1134	-1,0687	-1,1204	-0,47	*	> 0,150
3	U001	-0,8628	-0,5305	0,0158	-0,4592	0,66	*	> 0,150
4	U325	-1,4493	-1,3215	-0,6623	-0,9236	1,46	*	> 0,150
5	U385	-0,8709	-0,4621	-0,3058	-0,5463	-1,06	*	> 0,150
6	U447	-0,7124	-0,6218	-0,0870	-0,4737	1,48	*	> 0,150
7	CA0426	-1,0557	-0,8208	-0,0495	-0,3519	1,27	*	> 0,150
8	CA0534	-0,7974	-0,7974	-0,6152	-0,7367	1,57	*	> 0,150
9	CA0702	-0,9262	-0,8219	-0,4988	-0,7490	1,14	*	> 0,150
10	CA0810	-0,8932	-0,6565	-0,4592	-0,6696	-0,21	*	> 0,150
11	CA1120	-0,7348	-0,7129	0,3356	-0,3707	1,71	*	0,111

\* Nilai kurtosis tidak teridentifikasi,  $\alpha = 0,05$

**Tabel 4.13.** Hasil Uji Jumlah Komponen *Mixture* dalam Data *Chickpea* dengan dengan Pengembangan Algoritma RJMCMC

No	ID Gen	Uji Jumlah Komponen <i>Mixture</i>
Fungsi Ketahanan:		
1	LS0024	2
2	LS0035	1
3	LS0616	2
4	LS0159	2
5	LS0162	2
6	LS0185	2
7	LS0612	1
8	LS0081	2
9	LS0752	2
10	LS0759	1
11	LS0688	2
12	U070	3
13	U273	2
14	U278	1
15	U017	1
Fungsi Energi:		
1	LS0132	1
2	LS0341	1
3	U001	3
4	U325	2
5	U385	2
6	U447	2
7	CA0426	2
8	CA0534	2
9	CA0702	2
10	CA0810	3
11	CA1120	2



*Halaman ini sengaja dikosongkan*

## BAB V

### ***BAYESIAN UNIMODAL MODEL DAN BAYESIAN MODEL AVERAGING UNTUK IDENTIFIKASI DATA CHICKPEA***

Dalam bab ini akan disajikan beberapa algoritma yang terkait dengan identifikasi perbedaan ekspresi gen tanaman *Chickpea* di Indonesia melalui pemodelan Bayesian unimodal normal (BUM-normal) dan pemodelan Bayesian Averaging normal unimodal (BMA-normal unimodal). Selain itu, beberapa algoritma yang telah dibangun diimplementasikan pada data simulasi dan data *Chickpea*.

#### **5.1. Pemodelan dan Algoritma untuk Bayesian Unimodal Model (BUM) Normal**

Pemodelan Bayesian normal unimodal merupakan sebuah pemodelan yang menggunakan pendekatan Bayesian dimana distribusi normal dari data hanya memiliki satu modus (normal unimodal). Dasar perhitungannya menggunakan konsep Bayesian dimana penentuan distribusi posterior parameter model berdasarkan perkalian dari fungsi *likelihood* data dengan distribusi priornya. Pemodelan BUM-normal dibentuk menggunakan algoritma MCMC pendekatan *Gibbs Sampler*. Tahapan-tahapan algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) telah dijabarkan secara lengkap di sub bab 2.5.5.

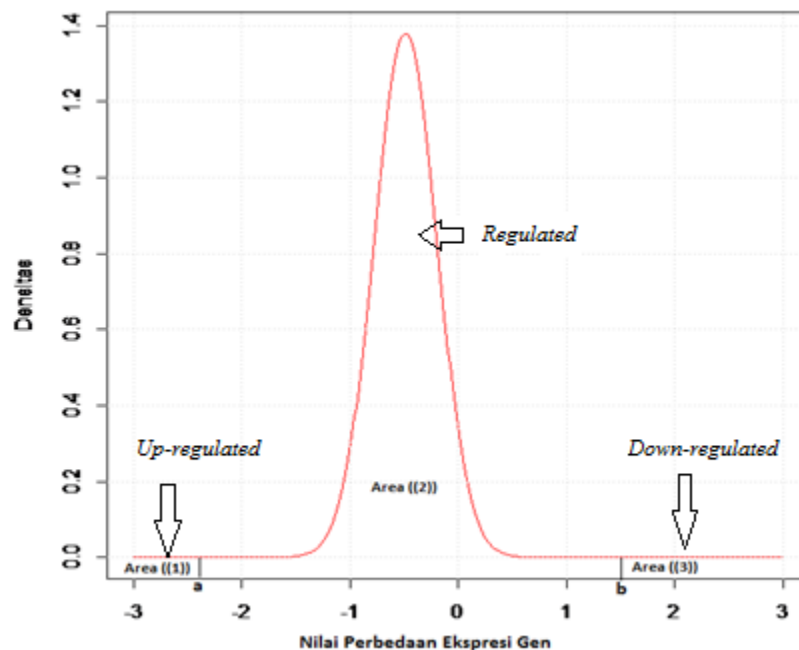
Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) merupakan sebuah algoritma dari metode MCMC yang menggunakan salah satu pendekatan umum untuk suatu  $\theta$  dengan  $g(\theta)$  yang berbentuk kompleks dengan metode *Gibbs Sampler*. Metode pendekatan ini merupakan suatu teknik simulasi untuk membangkitkan variabel random dari suatu distribusi tertentu secara tidak langsung tanpa harus menghitung densitasnya dan merupakan *generator* yang sangat efisien. Cara kerja metode *Gibbs Sampler* ini didasarkan pada sifat-sifat dasar dari *Markov Chain*, sehingga perhitungan yang sulit dapat dihindari karena konsep pentahapan proses *Markov* digunakan sebagai tahap iterasi dan proses estimasi parameter. Proses untuk membangkitkan variabel random dapat dilakukan dengan menerapkan langkah-langkah dari Algoritma 2.1.

Dalam hal ini  $\theta$  yang dimaksud adalah himpunan parameter yang memuat  $\lambda$ ,  $\sigma$  dan  $\tau$  sebagai parameter distribusi normal unimodal. Algoritma 2.1 akan menghasilkan distribusi setiap parameter dalam  $\theta$  yang dapat digunakan untuk penduga rata-rata bagi  $\lambda$ ,  $\sigma$  dan  $\tau$  dalam pemodelan BUM-normal, sehingga model tersebut dapat diestimasi. Selanjutnya, mengevaluasi pemenuhan asumsi pemodelan Bayesian pada model BUM-normal dengan kriteria *independence* antar data bangkitan numerik setiap parameter model, *irreducible*, *aperiodic* dan *recurrent* melalui plot fungsi autokorelasi, distribusi Kernel, *historydata* bangkitan numerik dan *quantiles* dari hasil program WinBUGS untuk mendapatkan model yang terbaik. Apabila model terbaik telah diperoleh, maka langkah selanjutnya adalah memvalidasi model BUM-normal dengan *goodness of fit* KS yang telah dijabarkan pada sub bab 2.7 dan menentukan nilai batas bawah dan batas atas HPD dengan *credible interval* 95% dari model BUM-normal terbaik sebagaimana yang telah dijabarkan pada sub bab 2.8. Langkah terakhir, menentukan kelompok data ke dalam group kategori *Down-regulated*, *Regulated*, dan *Up-regulated*.

Pemodelan BUM-normal dibentuk dari data yang memiliki distribusi normal unimodal dengan fungsi *likelihood* dari data sebagaimana telah dijabarkan dalam persamaan (2.28). Pemodelan dengan pendekatan Bayesian ini harus ditentukan distribusi prior setiap parameter dalam model. Prior dapat ditentukan dengan pendekatan struktur hirarki sebagaimana telah dijelaskan dalam sub bab 2.3.4. Langkah selanjutnya membentuk distribusi posterior setiap parameter yang merupakan kombinasi dari fungsi *likelihood* dan distribusi prior. Bentuk distribusi posterior yang diperoleh akan merupakan distribusi gabungan dari semua parameter dalam model BUM-normal sebagaimana yang telah dijabarkan dalam persamaan (2.29).

Distribusi bersyarat penuh dari masing-masing parameter BUM-normal yang menyusun distribusi posteriornya telah dijabarkan secara lengkap pada sub bab 2.5. Untuk memudahkan estimasi parameter model BUM-normal, struktur hirarki dari distribusi posterior digambarkan dalam DAG sebagaimana dalam Gambar 2.2. Tahapan-tahapan secara lengkap pemodelan BUM-normal akan disajikan dalam Algoritma 5.1.

Apabila model BUM-normal terbaik telah dibentuk, langkah selanjutnya membentuk HPD dengan *credible interval* 95% untuk densitas,  $f(\mathbf{x}|\boldsymbol{\theta})$ , BUM-normal, menggunakan konsep distribusi setimbang HPD yang proses lengkapnya telah dijabarkan dalam Algoritma 2.3. Pembentukan HPD ini dimaksudkan untuk mengelompokkan ID gen yang diamati ke dalam kelompok *Down-regulated*, *Regulated*, atau *Up-regulated*. Dalam HPD dengan *credible interval* 95%, akan ada tiga kelompok area yang ditetapkan, yaitu (1) Area-area yang memiliki nilai-nilai yang lebih kecil dari batas bawah namakan a, (2) Area-area yang memiliki nilai-nilai di antara nilai batas bawah namakan a dan batas atas namakan b, dan (3) Area-area yang memiliki nilai-nilai yang lebih besar dari batas atas namakan b. Ketiga area itu yang dapat mengelompokkan ID gen ke dalam kelompok *Up-regulated* di area (1), *Regulated* di area (2), dan *Down-regulated* di area (3). Sebagai contoh ilustrasi pembentukan HPD dengan *credible interval* 95% untuk model BUM-normal ditunjukkan pada Gambar 5.1.



**Gambar 5.1.** Ilustrasi Pembentukan HPD dengan *Credible Interval* 95% pada Model BUM-normal

Proses penetapan nilai-nilai batas dari HPD dengan *credible interval* 95% melalui langkah-langkah dalam Algoritma 2.3 sebagaimana yang telah dijabarkan dalam sub bab 2.8.

### Algoritma 5.1. Algoritma BUM-normal

Algoritma BUM-normal dibangun untuk memudahkan penetapan model BUM-normal dari data yang diamati, yaitu data berdistribusi normal unimodal sebagaimana yang telah dibahas pada Bab II. Dalam membentuk model BUM-normal digunakan bantuan program WinBUGS untuk mendapatkan penduga-penduga parameter model, baik  $\lambda$ ,  $\mu$  dan  $\sigma$ , dimana model BUM-normal dibangkitkan sebanyak 1.000 model dengan masing-masing parameter yang dibangkitkan sebanyak 5.000 kali. Langkah selanjutnya, mengevaluasi kebaikan model menurut kriteria pemenuhan pemodelan Bayesian melalui fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diduga (setiap deretan bangkitan parameter harus memenuhi kriteria *independence*, *irreducible*, *aperiodic*, dan *recurrent*). Model BUM-normal yang telah memenuhi kriteria *independence* antar iterasinya ditunjukkan oleh plot fungsi autokorelasi yang memiliki ciri hanya lag-0 yang signifikan berbeda dengan nol, sedangkan lag-lag selain lag-0 tidak signifikan berbeda dengan dengan nol dan konvergensi estimasi setiap parameter model telah memenuhi syarat, yaitu ditunjukkan oleh distribusi Kernel, *history*, dan *quantiles* yang berciri *irreducible*, *aperiodic* dan *recurrent*. Dalam langkah ini, akan diperoleh model BUM-normal terbaik, sehingga langkah berikutnya adalah mengevaluasi kebaikan model dengan *goodness of fit* KS. Pada langkah ini, konsep jarak terbesar antara densitas kumulatif empiris dan densitas kumulatif dugaan digunakan untuk mengetahui sejauh mana kebaikan model BUM-normal yang telah dibangun. Langkah terakhir adalah menetapkan batas-batas pengelompokan data melalui HPD dengan *credible interval* 95%. Melalui proses ini akan diketahui kelompok-kelompok ID gen yang diamati pada tiga kelompok, yaitu kelompok *Down-regulated*, *Regulated*, dan *Up-regulated*. Secara lengkap, proses penetapan model BUM-normal melalui langkah-langkah Algoritma 5.1.

Langkah 1. Bentuk model BUM-normal rata-rata dari 1.000 model yang dibangkitkan dengan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) dengan jumlah iterasi setiap parameter yang diduga dalam model sebanyak 5.000 kali.

Langkah 2. Evaluasi pemenuhan asumsi pemodelan Bayesian (*independence*, *irreducible*, *aperiodic*, dan *recurrent*) dari model BUM-normal rata-

rata yang terbaik berdasarkan hasil Langkah 1 melalui kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diduga dalam model.

- Langkah 3. Hitung nilai *goodness of fit* KS untuk validasi model sebagaimana dijabarkan dalam sub bab 4.6.
- Langkah 4. Tentukan nilai batas bawah (namakan a) dan batas atas (namakan b) HPD dengan *Credible Interval* 95% sebagaimana dijabarkan dalam sub 2.9.
- Langkah 5. Tetapkan kelompok data ke dalam tiga kelompok, yaitu *Down-regulated*, *Regulated*, dan *Up-regulated*.

## **5.2. Pemodelan dan Algoritma untuk Bayesian Model Averaging (BMA) Normal Unimodal**

Pemodelan BMA merupakan salah satu pemodelan pendekatan Bayesian yang menggunakan prinsip membentuk model tunggal terbaik berdasarkan semua kemungkinan model yang bisa dibentuk dan merata-ratakan distribusi posterior dari semua model terbaiknya untuk mendapatkan estimasi parameternya. Diharapkan pendekatan ini dapat memberikan estimasi model yang lebih tepat dibandingkan dengan prinsip pembentukan model tunggal yang mengabaikan adanya ketidakpastian model. Sebagaimana model BUM-normal, dasar perhitungan model BMA-normal unimodal juga menggunakan konsep Bayesian. Perbedaannya adalah distribusi posterior pada pemodelan BMA-normal unimodal ini merupakan hasil rata-rata distribusi posterior dari semua model terbaiknya.

Prediksi parameter dengan BMA menggunakan struktur model yang disusun sebagai kombinasi hirarki. Konsep yang digunakan dalam mengestimasi parameter dalam pemodelan BMA sebagaimana telah dijabarkan secara lengkap pada sub bab 2.6. Dalam pemodelan BMA-normal membutuhkan metode seleksi Occam's Window untuk menyeleksi model-model yang baik yang telah dibangkitkan. Berikut ini dijelaskan tentang metode seleksi Occam's Window sebagaimana yang dituangkan dalam Algoritma 5.2.

**Algoritma 5.2.** Algoritma Seleksi Model Occam's Window

Algoritma Seleksi Model Occam's Window digunakan untuk menyeleksi 1.000 model BUM-normal terbaik yang telah diperoleh berdasarkan Algoritma 2.1 untuk membentuk model BMA-normal unimodal terbaik. Proses seleksi model dengan metode Occam's Window melalui langkah-langkah berikut:

Langkah 1. Hitung nilai BIC (*Bayesian Information Criterion*) dan maxBIC (nilai BIC terbesar) untuk 1.000 model BUM-normal terbaik, dimana rumus BIC sebagaimana telah dijabarkan dalam persamaan (2.41).

Langkah 2. Hitung nilai peluang posterior ( $P(M_k | X)$ ) dan Max  $P(M_k | X)$  dari masing-masing model dengan rumus peluang posterior ( $P(M_k | X)$ ) untuk setiap model adalah sebagai berikut:

$$P(M_k | X) = \frac{e^{-0,5*(BIC - \max BIC)}}{\sum_{i=1}^m e^{-0,5*(BIC - \max BIC)}}.$$

Langkah 3. Hitung nilai  $\mathcal{A}'$  untuk kriteria seleksi model BMA-normal unimodal terbaik dengan rumus sebagai berikut:

$$\mathcal{A}' = \{M_k : \frac{\max_l \{P(M_k | X)\}}{P(M_k | X)} \leq c = 20\} \propto \alpha = 5\%.$$

Langkah 4. Tetapkan hasil seleksi berdasarkan Langkah 3. Jika nilai  $\mathcal{A}' \leq 20$ , maka model ke- $k$  sesuai untuk data dan masuk dalam pemodelan BMA-normal unimodal. Sebaliknya jika nilai  $\mathcal{A}' > 20$ , maka model ke- $k$  tidak sesuai untuk data dan tidak diikutkan dalam pemodelan BMA-normal unimodal (Madigan dan Raftery, 1994).

### Algoritma 5.3. Algoritma BMA-Normal Unimodal

Algoritma BMA-normal unimodal digunakan untuk membentuk model BMA-normal unimodal terbaik sehingga pengelompokan data ke dalam *Down-regulated*, *Regulated*, dan *Up-regulated* dapat diperoleh. Proses pemodelan BMA-normal unimodal dapat dilakukan melalui langkah-langkah berikut:

- Langkah 1. Bentuk 1.000 model BUM-normal terbaik dengan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) dengan jumlah iterasi parameter yang akan diduga dalam model sebanyak 5.000 kali.
- Langkah 2. Evaluasi pemenuhan asumsi pemodelan Bayesian (*independence*, *irreducible*, *aperiodic*, dan *recurrent*) dari 1.000 model BUM-normal rata-rata yang terbaik berdasarkan hasil Langkah 1 melalui kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diduga dalam model
- Langkah 3. Seleksi model dengan metode seleksi Occam's Window untuk membentuk model BMA-normal unimodal dengan menerapkan Algoritma 5.2
- Langkah 4. Hitung nilai *goodness of fit* KS untuk validasi model sebagaimana yang telah dijabarkan dalam sub bab 4.6.
- Langkah 5. Tentukan nilai batas bawah (namakan a) dan batas atas (namakan b) HPD dengan *credible interval* 95% sebagaimana telah dijabarkan dalam sub bab 2.9.
- Langkah 6. Tetapkan kelompok data ke dalam tiga kelompok, yaitu *Down-regulated*, *Regulated*, dan *Up-regulated*.

Pada sub bab berikutnya akan disajikan hasil implementasi dari semua algoritma yang telah dibangun untuk membentuk model BUM-normal maupun model BMA-normal unimodal pada data simulasi maupun data *Chickpea*. Hasil implementasi dari semua algoritma yang telah dibangun untuk pemodelan BMA-normal unimodal (Astuti dkk., (2014a; 2014b; 2015b)).



### 5.3. Implementasi Pemodelan untuk Data Simulasi dan Data *Chickpea*

Implementasi pemodelan untuk data simulasi dan data *Chickpea* dilakukan untuk membentuk model BUM-normal dan model BMA-normal unimodal di kedua data tersebut. Proses pembentukan modelnya dengan mengimplementasi Algoritma 5.1 sampai dengan Algoritma 5.3.

#### 5.3.1. Pemodelan BUM-Normal Data Simulasi

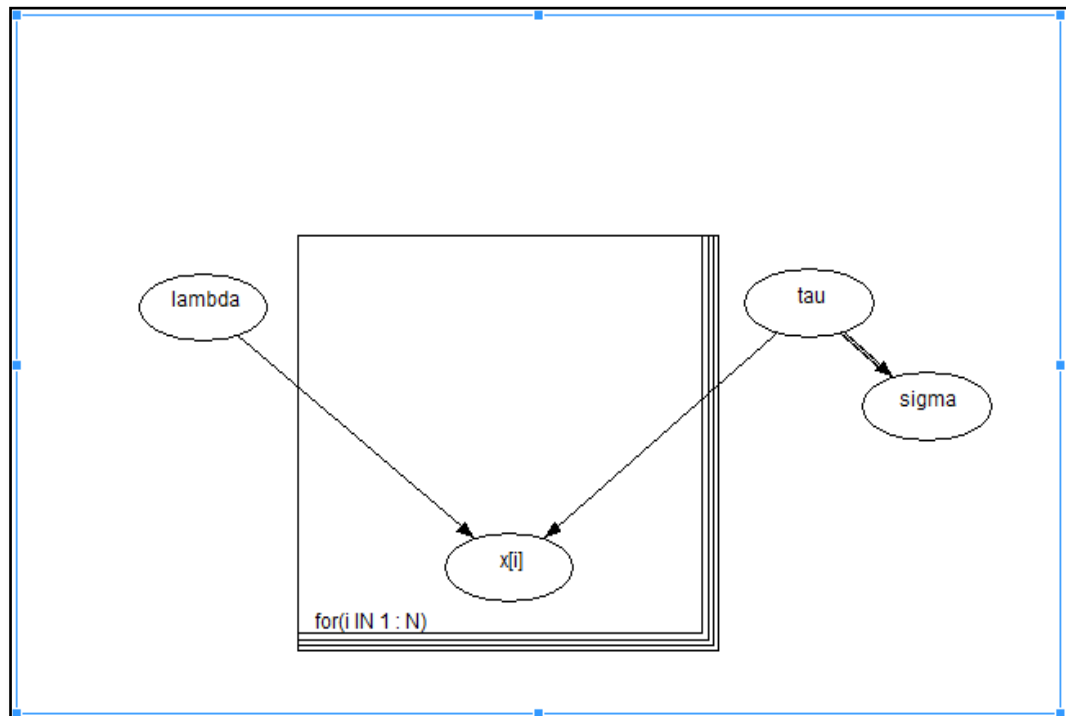
Pemodelan BUM-Normal akan dibentuk dengan tahapan-tahapan sebagai berikut:

##### a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian

Berdasarkan data simulasi yang telah dibangun dengan ciri data *Chickpea* sebagaimana yang telah dijelaskan dalam sub bab 4.4, teridentifikasi lima ID gen dari fungsi ketahanan yang memiliki distribusi normal unimodal, yaitu ID gen: LS0035, LS0612, LS0759, U278, dan U017. Sedangkan ID gen dari fungsi energi, hanya ada dua ID gen yang memiliki distribusi normal unimodal yaitu ID gen: LS0132 dan LS0341. Hal ini berarti dari 26 ID gen yang diamati, ada tujuh ID gen yang memiliki distribusi normal unimodal sehingga data tujuh ID gen ini yang akan dimodelkan sebagai bentuk normal unimodal dengan pendekatan Bayesian (BUM-normal).

Estimasi parameter untuk model BUM-normal dilakukan dengan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1). Implementasi dari Algoritma ini menggunakan Program WinBUGS 1.4 dengan membuat *Graphical Model* (*Doodle BUGS*) (Gambar 5.2) atau dengan membuat *Syntax BUGS* (*Syntax 5.1*) untuk tujuh ID gen yang telah teridentifikasi memiliki distribusi normal unimodal. Sebagai contoh diberikan *Doodle/DAG BUGS* (Gambar 5.2) dan *Syntax BUGS* (*Syntax 5.1*) untuk data ID gen LS0035.

Berdasarkan *running* program BUGS, dalam hal ini sebagai contoh diberikan untuk ID gen LS0035 dan hasilnya disajikan pada Tabel 5.1, Gambar 5.3 sampai dengan Gambar 5.6. Hasil yang disajikan merupakan rata-rata dari pembangkitan 5.000 kali untuk setiap parameter yang akan diestimasi dan dibangkitkan 1.000 model data simulasi untuk ID Gen LS0035. Hasil secara lengkap untuk enam ID Gen yang lain, yaitu LS0612, LS0759, U278, U017, LS0132, dan LS0341 dapat dilihat di Lampiran 3.



**Gambar 5.2.** Graphical Model (Doodle BUGS) untuk Model BUM-Normal

**Syntax 5.1.** SyntaxBUGS untuk Model BUM-Normal

```

model;
{
  for( i in 1 : N ) {
    x[i] ~ dnorm(lambda,tau)
  }
  tau ~ dgamma(11.28,1)
  sigma <- 1 / tau
  lambda ~ dnorm( 0.1,0.05)
}

#1.Data Simulasi: LS0035
DATA
list(x=c(-0.6251,-0.4767,-0.3778),N=3)

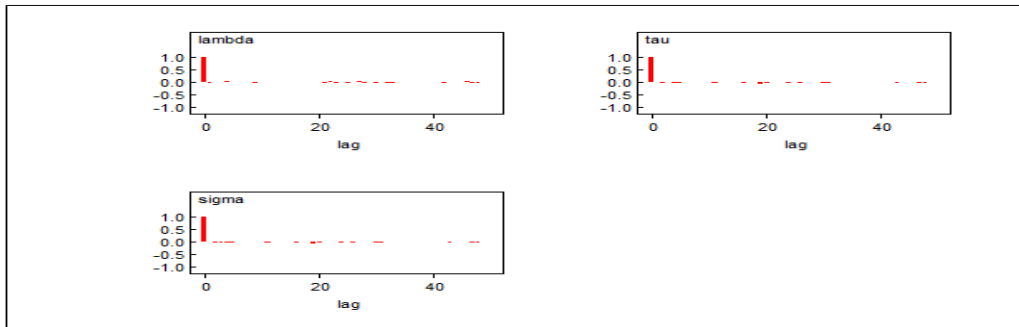
#1.Data Chiekpea: LS0035
DATA
list(x=c(-0.680046,-0.324557,-0.473301),N=3)

INITIAL
list(lambda=1)

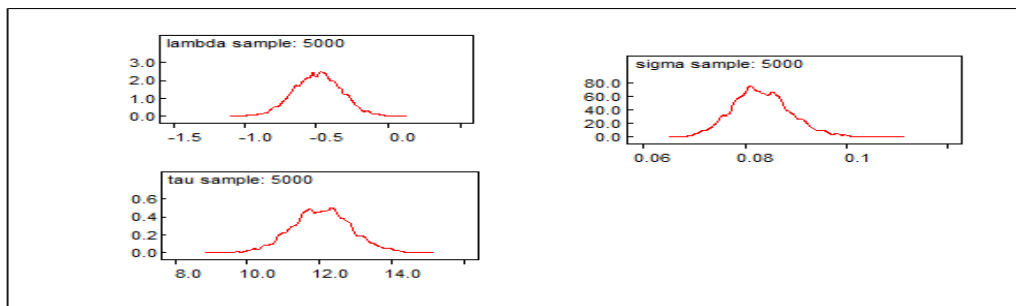
```

**Tabel 5.1.** Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen LS0035

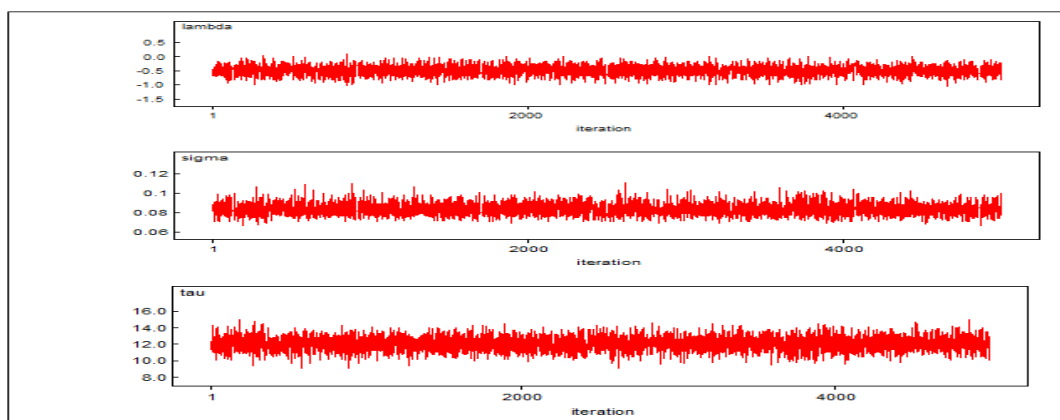
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,493	0,1736	0,0021	-0,8293	-0,4937	-0,1534	1	5.000
sigma	0,0902	0,0279	4,23E-01	0,0505	0,0857	0,1608	1	5.000
tau	12,06	3,4570	0,0519	6,2260	11,670	19,90	1	5.000



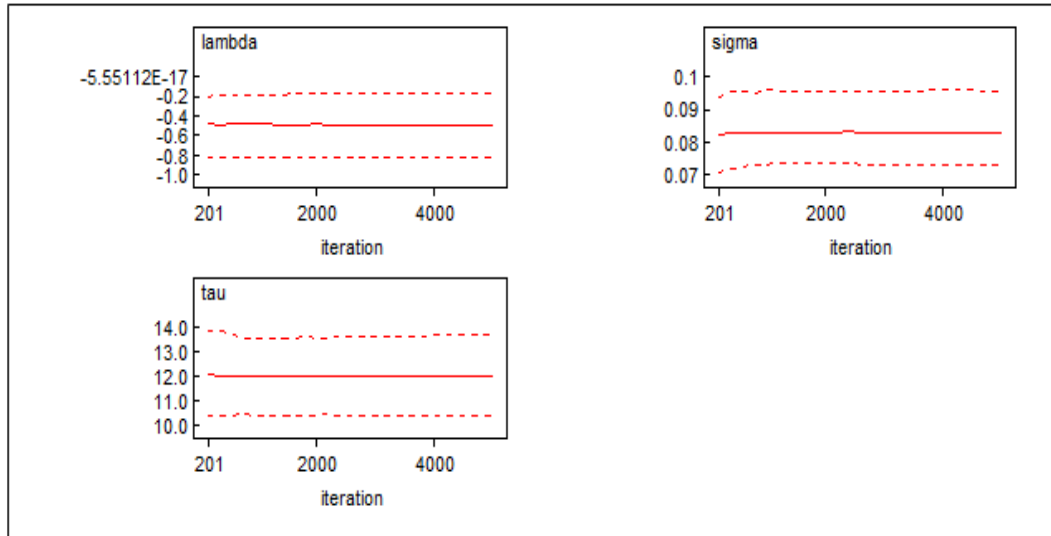
**Gambar 5.3.** Fungsi Autokorelasi setiap Parameter yang di Estimasi untuk ID Gen LS0035



**Gambar 5.4.** Distribusi Kernel setiap Parameter yang di Estimasi untuk ID Gen LS0035



**Gambar 5.5.** History Data Bangkitan setiap Parameter yang di Estimasi untuk ID Gen LS0035



**Gambar 5.6.** *Quantiles* setiap Parameter yang di Estimasi untuk ID Gen LS0035

Berdasarkan hasil evaluasi estimasi parameter model BUM-normal pada Tabel 5.1 dan Gambar 5.3 sampai dengan Gambar 5.6 serta Lampiran 3 dapat diketahui bahwa untuk data simulasi tujuh ID gen, yaitu lima ID gen dengan fungsi ketahanan: LS0035, LS0612, LS0759, U278, dan U017 serta dua ID gen dengan fungsi energi: LS0132 dan LS0341 yang telah dibentuk sudah sangat baik. Hal ini dapat diketahui dari plot fungsi autokorelasi semua parameter yang diestimasi (Gambar 5.3) mempunyai nilai nyata berbeda dengan nol untuk lag-0 dan lag-lag selain lag-0 mempunyai nilai tidak nyata berbeda dengan nol. Hal ini menunjukkan bahwa sifat *independence* parameter antar iterasi dalam sampel MCMC sudah terpenuhi. Demikian juga dengan plot distribusi Kernel, *history*, dan *quantiles* dari setiap parameter yang diestimasi (Gambar 5.4 sampai dengan Gambar 5.6) dimana konvergensi dari estimasi setiap parameter telah terpenuhi. Hal ini ditunjukkan dengan terpenuhinya sifat *irreducible*, *aperodic* dan *recurrent*. Oleh karena itu telah diperoleh model BUM-normal terbaik untuk tujuh ID gen tersebut.

#### **b. Validasi Model BUM-Normal**

Untuk memvalidasi model BUM-normal yang telah terbentuk digunakan *goodness of fit* Kolmogorov-Smirnov (KS). Apabila nilai KS kecil, yaitu mendekati nilai nol, maka model yang divalidasi makin baik. Hasil selengkapnya

untuk tujuh ID gen, yaitu LS0035, LS0612, LS0759, U278, U017, LS0132, dan LS0341 disajikan dalam Tabel 5.2. Berdasarkan hasil validasi model data simulasi ini dapat diketahui bahwa model BUM-normal mampu memodelkan tujuh ID gen yang diamati dengan baik. Hal ini dapat diketahui dari nilai rata-rata KS model BUM-normal secara umum memiliki nilai yang cukup kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$ . Tetapi model OLS sebagaimana yang telah dijabarkan dalam sub bab 2.2, memiliki nilai rata-rata KS yang secara umum lebih kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$  dibandingkan nilai rata-rata KS model BUM-normal.

**Tabel 5.2.** *Goodness of Fit* KS Model BUM-Normal dan Model OLS untuk Data Simulasi

No	ID Gen	Nilai Rata-Rata Kolmogorov-Smirnov (KS) dari Pembangkitan Data 1.000 Kali	
		Model OLS	BUM-Normal
Fungsi Ketahanan:			
1	LS0035	0,2710	0,3589
2	LS0612	0,2820	0,4768
3	LS0759	0,2841	0,4810
4	U278	0,2810	0,3540
5	U017	0,2803	0,4332
Fungsi Energi:			
6	LS0132	0,2819	0,4436
7	LS0341	0,2874	0,4307

$$KS_{tabel} = 0,708$$

### c. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen

Pembentukan HPD dengan *credible interval* 95% dilakukan untuk menentukan nilai-nilai batas *credible interval* dari model yang dibangun, sehingga dapat diketahui pengelompokan dari tujuh ID gen yang diamati ke dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% diperoleh dengan mengimplementasi Algoritma 2.3. Hasil selengkapnya disajikan dalam Tabel 5.3.

**Tabel 5.3.** Nilai Batas HPD dengan *Credible Interval* 95% Model BUM-Normal untuk Data Simulasi

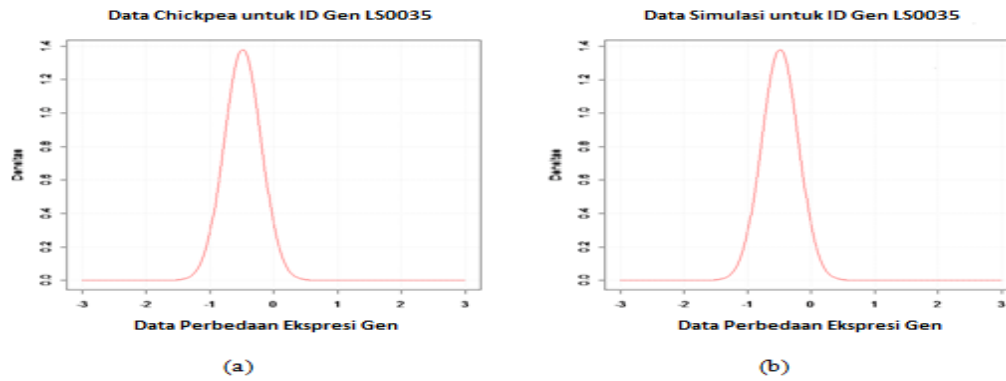
No	ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BUM-Normal Data Simulasi		Hasil Pengelompokan
		Batas Bawah	Batas Atas	
Fungsi Ketahanan:				
1	LS0035	-1,0586	0,0730	<i>Regulated</i>
2	LS0612	-1,7719	-0,6396	<i>Up-regulated</i>
3	LS0759	-1,3408	-0,2061	<i>Up-regulated</i>
4	U278	-1,0863	0,0476	<i>Regulated</i>
5	U017	-0,8254	0,3065	<i>Regulated</i>
Fungsi Energi:				
6	LS0132	-1,0867	0,0460	<i>Regulated</i>
7	LS0341	-1,6781	-0,5464	<i>Up-regulated</i>

Berdasarkan Tabel 5.3 dapat diketahui bahwa ada tiga ID gen yang masuk group *Up-regulated*, yaitu dua ID gen dari fungsi ketahanan, yaitu ID gen LS0612 dan LS0759 serta satu ID gen dari fungsi energi, yaitu ID gen LS0341. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas yang bernilai negatif (-). Sedangkan empat ID gen yang lain masuk group *Regulated*, yaitu tiga ID gen dengan fungsi ketahanan: LS0035, U278, dan U017 serta satu ID gen dengan fungsi energi, yaitu ID gen LS0132. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Hasil yang diperoleh dengan pemodelan BUM-normal teridentifikasi tiga ID Gen, yaitu ID gen LS0612, LS0759, dan LS0341 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia.

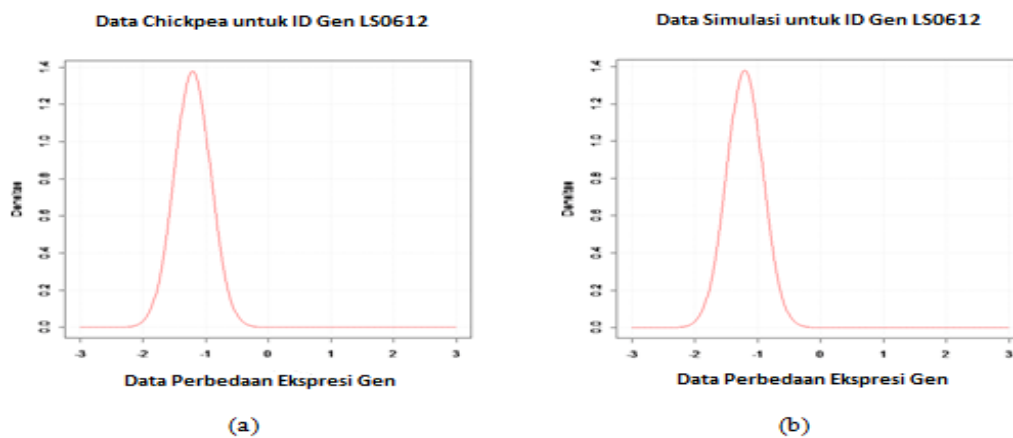
#### **d. Bentuk Distribusi Model BUM-Normal**

Data simulasi dibangun berdasarkan pola dari data *Chickpea*. Oleh karena itu, bentuk distribusi model BUM-normal disajikan secara bersama-sama dengan bentuk distribusi model BUM-normal data *Chickpea*. Bentuk distribusi model BUM-normal data simulasi dan data *Chickpea* untuk tujuh ID gen disajikan dalam Gambar 5.7 sampai dengan Gambar 5.13. Berdasarkan hasil distribusi yang diperoleh dapat diketahui bahwa distribusi data simulasi mengikuti distribusi data

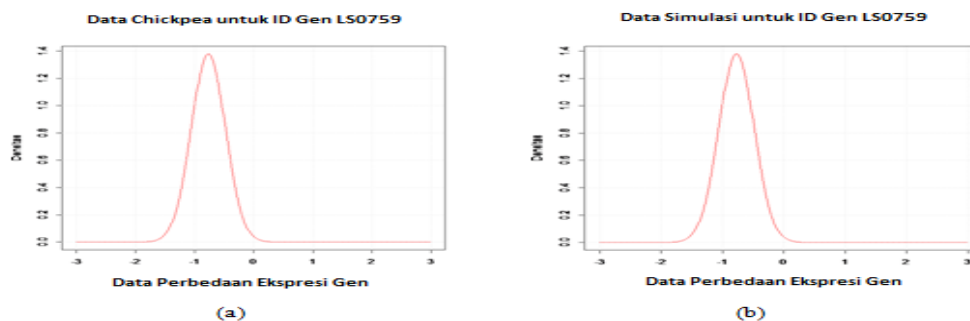
*Chickpea*. Selain itu, dari tujuh ID gen yang diamati memiliki distribusi normal unimodal yang sama tetapi berbeda-beda pada ciri nilai parameter model yang diperoleh.



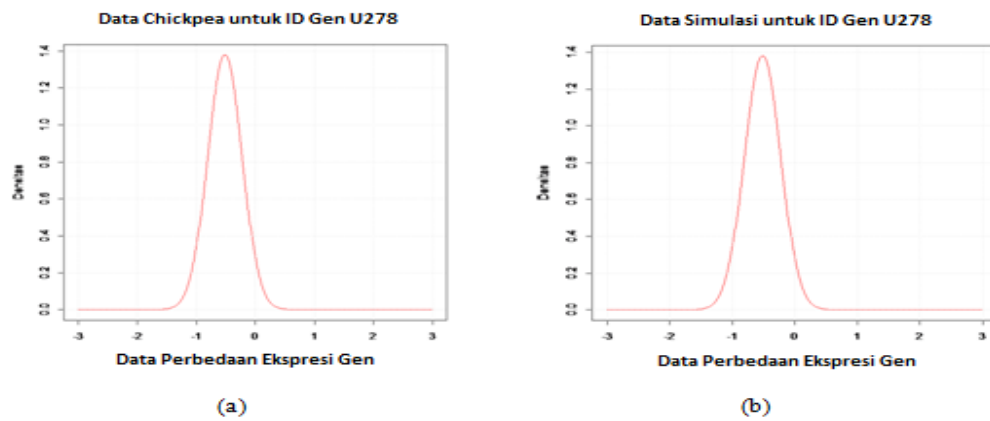
**Gambar 5.7.** Bentuk Distribusi BUM-normal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0035



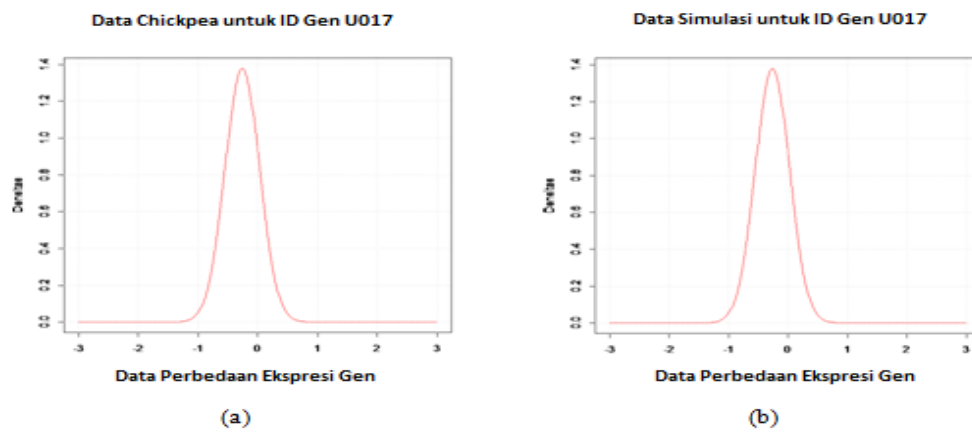
**Gambar 5.8.** Bentuk Distribusi BUM-normal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0612



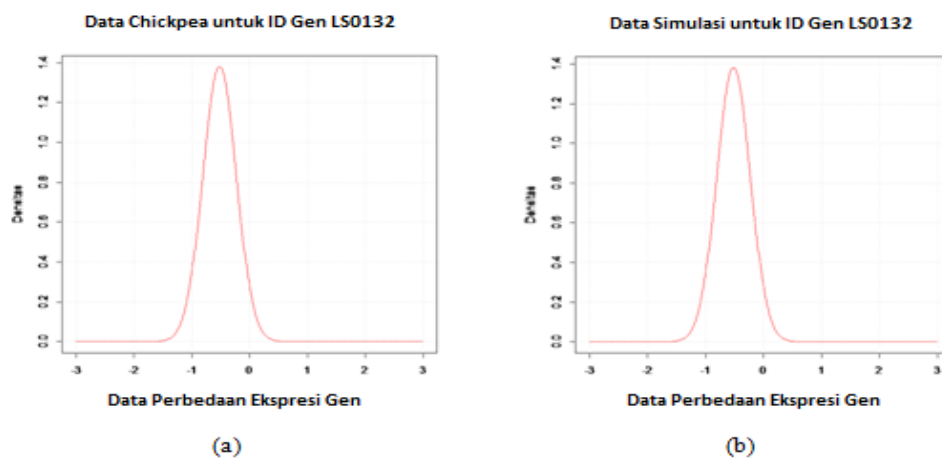
**Gambar 5.9.** Bentuk Distribusi BUM-normal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0759



**Gambar 5.10.** Bentuk Distribusi BUM-normal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U278

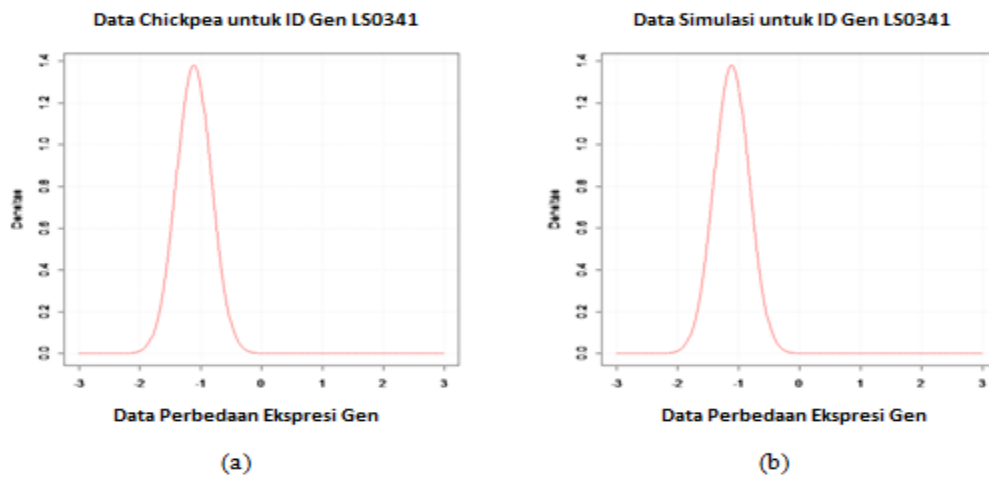


**Gambar 5.11.** Bentuk Distribusi BUM-normal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U017



**Gambar 5.12.** Bentuk Distribusi BUM-normal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0132





**Gambar 5.13.** Bentuk Distribusi BUM-normal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0341

### 5.3.2. Pemodelan BMA-Normal Unimodal Data Simulasi

Pemodelan BMA-normal unimodal untuk data simulasi melalui tahapan-tahapan sebagai berikut:

#### a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian

Implementasi darisemua langkah dari Algoritma 2.1 dan Langkah 1 dari Algoritma 5.3 untuk tujuh ID gen yang diamati, yaitu LS0035, LS0612, LS0759, U278, U017, LS0132, dan LS0341, disajikan secara lengkap di Lampiran 3. Berdasarkan hasil evaluasi rata-rata dari 5.000 pembangkitan parameter dan 1.000 model yang telah dibentuk dapat diketahui bahwa model tersebut sudah sangat baik dengan terpenuhinya sifat *independence*, *irreducible*, *aperiodic* dan *recurrent* sebagaimana yang telah ditunjukkan pada Gambar 5.3 sampai dengan Gambar 5.6 serta gambar-gambar pada Lampiran 3. Oleh karena itu, proses ini telah berhasil memperoleh model BUM-normal terbaik untuk tujuh ID gen tersebut.

#### b. Seleksi Model dengan Metode Occam's Window

Langkah 3 dari Algoritma 5.3 akan melakukan seleksi model untuk 1.000 model BUM-normal yang telah dibangun. Seleksi untuk 1.000 model BUM-normal terbaik dilakukan menggunakan metode seleksi Occam's Window dengan

mengimplementasikan Algoritma 5.2 agar dapat diperoleh model BMA-normal unimodal terbaik. Hasil yang diperoleh disajikan secara lengkap pada Tabel 5.4.

Berdasarkan hasil seleksi Occam's Window pada Tabel 5.4 dapat diketahui bahwa pada pemodelan BMA-normal unimodal yang dibangun melibatkan hampir seluruh model yang dibangkitkan, yaitu rata-rata ada 981 model yang terseleksi dan membangun model BMA-normal unimodal dari masing-masing tujuh ID gen yang diamati. Model BMA-normal unimodal yang terbangun cukup kompleks, tetapi dengan kemajuan teknologi komputasi yang semakin maju dan berkembang saat ini, maka kompleksitas model bukan menjadi masalah dalam estimasi modelnya.

**Tabel 5.4.** Hasil Seleksi Model dengan Metode Occam's Window pada Pemodelan BMA-Normal Unimodal untuk Data Simulasi

No	ID Gen	Hasil Seleksi 1.000 Model BUM-Normal dengan Metode Occam's Window dari Pengulangan sebanyak 1.000 kali untuk setiap ID Gen
Fungsi Ketahanan:		
1	LS0035	981,24
2	LS0612	980,97
3	LS0759	980,93
4	U278	981,29
5	U017	980,67
Fungsi Energi:		
6	LS0132	980,71
7	LS0341	980,65
Rata-Rata		980,92

### c. Validasi Model BMA-Normal Unimodal

Untuk memvalidasi model BMA-normal unimodal yang telah terbentuk digunakan *goodness of fit* KS. Hasil selengkapnya untuk tujuh ID gen, yaitu LS0035, LS0612, LS0759, U278, U017, LS0132, dan LS0341 disajikan dalam Tabel 5.5.

Berdasarkan Tabel 5.5 dapat diketahui bahwa model BMA-normal unimodal mampu memodelkan tujuh ID gen yang diamati dengan baik. Hal ini dapat diketahui dari nilai rata-rata KS model BMA-normal unimodal secara

umum memiliki nilai cukup kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$ . Untuk model OLS sebagaimana yang telah dijabarkan dalam sub bab 2.2, memiliki nilai rata-rata KS yang secara umum lebih kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$  dibandingkan dengan nilai rata-rata KS model BMA-normal unimodal.

#### **d. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen**

Pembentukan HPD dengan *credible interval* 95% dilakukan untuk menentukan nilai-nilai batas *credible interval* dari model yang telah dibangun untuk mengetahui pengelompokan tujuh ID gen yang diamati ke dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% diperoleh dengan mengimplementasi Algoritma 2.3. Hasil selengkapnya disajikan dalam Tabel 5.6.

Berdasarkan Tabel 5.6 dapat diketahui bahwa ada tiga ID gen yang masuk group *Up-regulated*, yaitu dua ID gen dari fungsi ketahanan: ID gen LS0612 dan LS0759 serta satu ID gen dari fungsi energi: ID gen LS0341. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai negatif (-). Sedangkan empat ID gen yang lain masuk group *Regulated*, yaitu tiga ID gen dengan fungsi ketahanan: LS0035, U278, dan U017 serta satu ID gen dengan fungsi energi: ID gen LS0132. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Hasil yang diperoleh dengan pemodelan BMA-normal unimodal teridentifikasi tiga ID Gen: LS0612, LS0759, dan LS0341 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia.

### 5.3.3. Pemodelan BUM-Normal untuk Data *Chickpea*

Pemodelan BUM-normal untuk data *Chickpea* dilakukan untuk mendapatkan model BUM-normal bagi kasus aplikasi data *microarray* di Indonesia yaitu data *Chickpea* sebagai hasil dari penelitian Harijati (2007). Tahapan-tahapan pemodelan BUM-normal dijabarkan sebagai berikut:

#### a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian

Semua algoritma yang telah dibangun untuk membentuk model BUM-normal diimplementasikan pada data *Chickpea*. Sebagai contoh akan disajikan hasil analisis ID Gen LS0035 dari Program WinBUGS untuk membentuk model BUM-normal (Tabel 5.7 dan Gambar 5.14 sampai dengan Gambar 5.17). Sedangkan hasil selengkapnya untuk tujuh ID gen yang diamati, lima ID gen dengan fungsi ketahanan dan dua ID gen dengan fungsi energi, yaitu ID gen: LS0035, LS0612, LS0759, U278, U017, LS0132, dan LS0341 disajikan pada Lampiran 3.

**Tabel 5.5.** *Goodness of Fits* KS Model BMA-Normal Unimodal dan Model OLS untuk Data Simulasi

No	ID Gen	Kolmogorov-Smirnov (KS)	
		Model OLS	BMA-Normal Unimodal
Fungsi Ketahanan:			
1	LS0035	0,2710	0,3672
2	LS0612	0,2820	0,4754
3	LS0759	0,2841	0,4795
4	U278	0,2810	0,3613
5	U017	0,2803	0,4362
Fungsi Energi:			
6	LS0132	0,2819	0,4464
7	LS0341	0,2874	0,4344

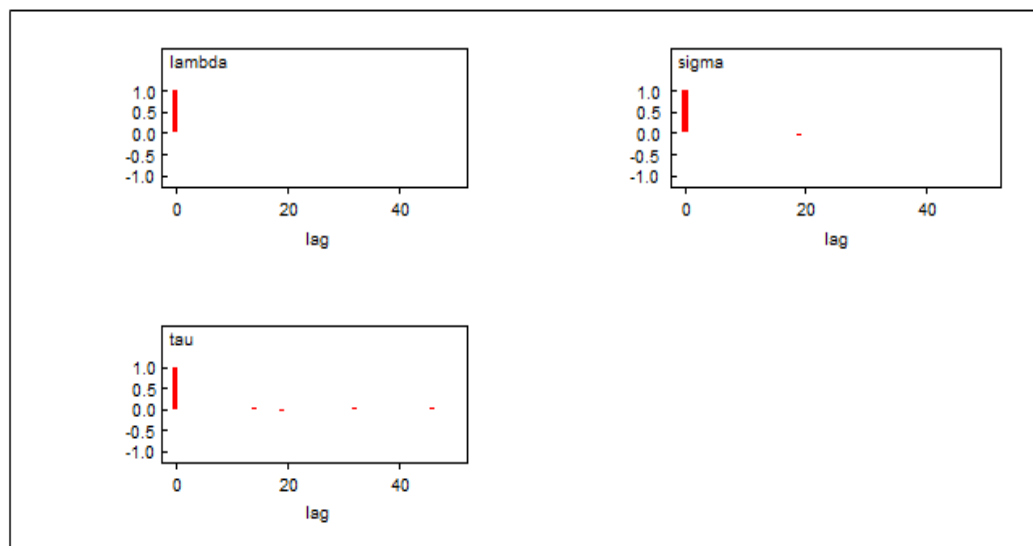
$$KS_{tabel} = 0,708$$

**Tabel 5.6.** HPD dengan *Credible Interval* 95% Model BMA-Normal Unimodal untuk Data Simulasi

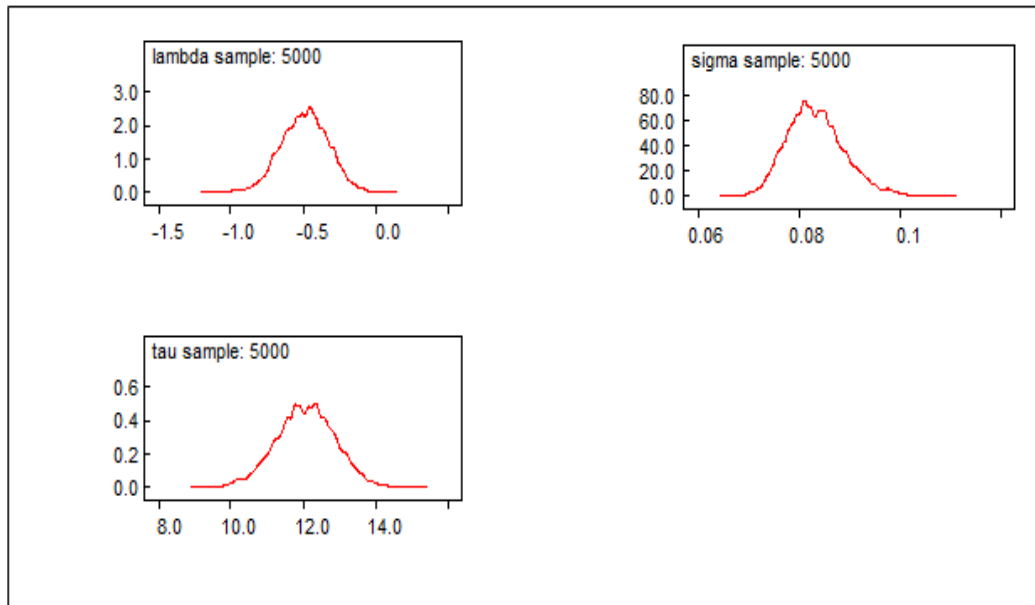
No	ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BMA-Normal Unimodal untuk Data Simulasi		Hasil Pengelompokan
		Batas Bawah	Batas Atas	
Fungsi Ketahanan:				
1	LS0035	-1,1011	0,1207	<i>Regulated</i>
2	LS0612	-1,8219	-0,6007	<i>Up-regulated</i>
3	LS0759	-1,3835	-0,1623	<i>Up-regulated</i>
4	U278	-1,1304	0,0916	<i>Regulated</i>
5	U017	-0,8758	0,3455	<i>Regulated</i>
Fungsi Energi:				
6	LS0132	-1,1311	0,0903	<i>Regulated</i>
7	LS0341	-1,7275	-0,5063	<i>Up-regulated</i>

**Tabel 5.7.** Estimasi Parameter Data *Chickpea* untuk ID Gen LS0035

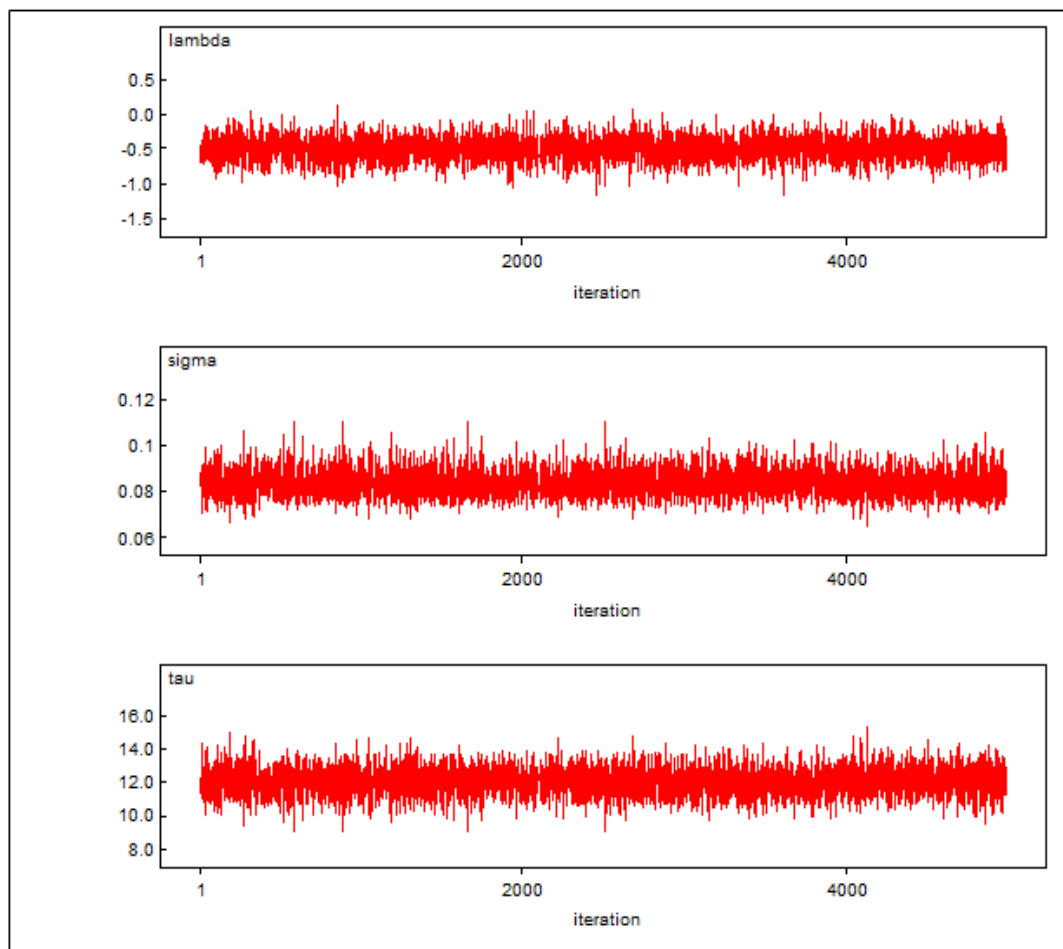
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,4924	0,1749	0,0021	-0,8315	-0,4931	-0,1501	1	5.000
Sigma	0,0917	0,0283	4,30E-01	0,0513	0,0871	0,1634	1	5.000
Tau	11.87	3,402	0,0511	6,127	11,49	19,58	1	5.000



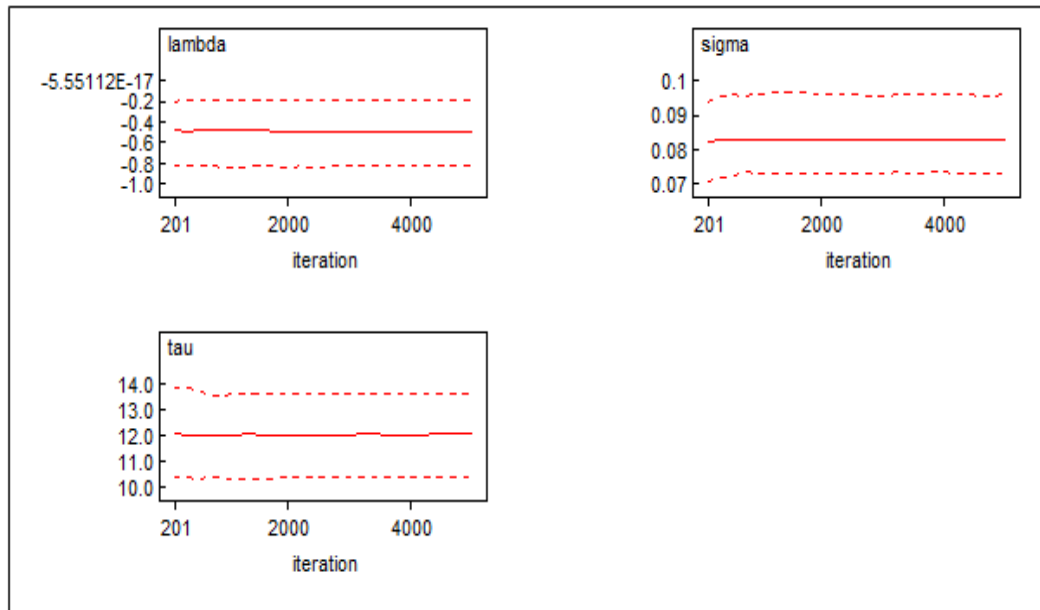
**Gambar 5.14.** Fungsi Autokorelasi setiap Parameter yang di Estimasi untuk ID Gen LS0035



**Gambar 5.15.** Distribusi Kernel setiap Parameter yang di Estimasi untuk ID Gen LS0035



**Gambar 5.16.** History Data Bangkitan setiap Parameter yang di Estimasi untuk ID Gen LS0035



**Gambar 5.17.** Quantiles setiap Parameter yang di Estimasi untuk ID Gen LS0035

Berdasarkan hasil evaluasi estimasi parameter model BUM-normal data *Chickpea* dalam Tabel 5.7 dan Gambar 5.14 sampai dengan Gambar 5.17 serta Lampiran 4 dapat diketahui bahwa untuk data tujuh ID gen, yaitu lima ID gen dengan fungsi ketahanan: LS0035, LS0612, LS0759, U278, dan U017 serta dua ID gen dengan fungsi energi: LS0132 dan LS0341 yang telah dibentuk sudah sangat baik. Hal ini dapat diketahui dari plot fungsi autokorelasi semua parameter yang diestimasi (Gambar 5.14) mempunyai nilai nyata berbeda dengan nol untuk lag-0 dan lag-lag selain lag-0 mempunyai nilai tidak nyata berbeda dengan nol. Hal ini menunjukkan bahwa sifat *independence* parameter antar iterasi dalam sampel MCMC sudah terpenuhi. Demikian juga dengan plot distribusi Kernel, *history*, dan *quantiles* dari setiap parameter yang diestimasi (Gambar 5.15 sampai dengan Gambar 5.17) dimana konvergensi dari estimasi setiap parameter telah terpenuhi. Hal ini ditunjukkan dengan terpenuhinya sifat *irreducible*, *aperodic* dan *recurrent*. Oleh karena itu telah diperoleh model BUM-normal terbaik untuk tujuh ID gen data *Chickpea*.

### b. Validasi Model BUM-Normal

Model BUM-normal untuk data *Chickpea* yang telah dibentuk untuk tujuh ID gen selanjutnya dilakukan validasi model dengan *goodness of fit* KS. Hasil selengkapnya disajikan dalam Tabel 5.8.

**Tabel 5.8.** *Goodness of Fits* KS Model BUM-Normal dan Model OLS untuk Data *Chickpea*

No	ID Gen	Kolmogorov-Smirnov (KS)	
		Model OLS	BUM-Normal
Fungsi Ketahanan:			
1	LS0035	0,2098	0,2878
2	LS0612	0,3021	0,4463
3	LS0759	0,2937	0,4723
4	U278	0,2528	0,2914
5	U017	0,2711	0,3845
Fungsi Energi:			
6	LS0132	0,2068	0,3711
7	LS0341	0,2087	0,3706

$$KS_{tabel} = 0,708$$

Berdasarkan Tabel 5.8 dapat diketahui bahwa model BUM-normal cukup baik dalam memodelkan tujuh ID gen yang diamati. Hal ini dapat diketahui dari nilai KS yang secara umum memiliki nilai yang cukup kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari nilai  $KS_{tabel} = 0,708$  untuk masing-masing ID gen. Untuk model OLS sebagaimana yang telah dijabarkan dalam sub bab 2.2, memiliki nilai KS secara umum lebih kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari nilai  $KS_{tabel} = 0,708$  dibandingkan dengan nilai KS model BUM-normal.

### c. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen

Pembentukan HPD dengan *credible interval* 95% untuk data *Chickpea* dilakukan untuk menentukan nilai-nilai batas *credible interval* dari model yang dibangun untuk mengetahui pengelompokan dari tujuh ID gen yang diamati ke dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% untuk data *Chickpea* diperoleh dengan



mengimplementasikan Algoritma 2.3. Hasil secara lengkap disajikan dalam Tabel 5.9.

**Tabel 5.9.** HPD dengan *Credible Interval* 95% Model BUM-Normal untuk Data *Chickpea*

No		ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BUM-Normal Data <i>Chickpea</i>		Hasil Pengelompokan
			Batas Bawah	Batas Atas	
Fungsi Ketahanan:					
1		LS0035	-1,2242	-0,0460	<i>Up-regulated</i>
2		LS0612	-1,9817	-0,8903	<i>Up-regulated</i>
3		LS0759	-1,2153	-0,1115	<i>Up-regulated</i>
4		U278	-0,8752	0,2848	<i>Regulated</i>
5		U017	-0,6038	0,5521	<i>Regulated</i>
Fungsi Energi:					
6		LS0132	-1,0872	0,0378	<i>Regulated</i>
7		LS0341	-1,9324	-0,8156	<i>Up-regulated</i>

Berdasarkan Tabel 5.9 dapat diketahui bahwa ada empat ID gen yang masuk group *Up-regulated*, yaitu tiga ID gen dari fungsi ketahanan, LS0035, LS0612 dan LS0759 serta satu ID gen dari fungsi energi, yaitu ID gen LS0341. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai negatif (-). Sedangkan tiga ID gen yang lain masuk group *Regulated*, yaitu dua ID gen dengan fungsi ketahanan: U278, dan U017 serta satu ID gen dengan fungsi energi: LS0132. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Sementara tidak satupun ditemukan ID gen yang masuk dalam kategori *Down-regulated*. Hasil yang diperoleh ini berarti dengan pemodelan BUM-normal untuk data *Chickpea* teridentifikasi empat ID Gen, yaitu LS0035, LS0612, LS0759, dan LS0341 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia.

#### 5.3.4. Pemodelan BMA-Normal Unimodal Data *Chickpea*

Sebagaimana pemodelan BUM-normal, pemodelan BMA-normal unimodal untuk data *Chickpea* juga dilakukan untuk mendapatkan model BMA-

normal unimodal bagi kasus aplikasi data *microarray* di Indonesia yaitu data *Chickpea* sebagai hasil dari penelitian Harijati (2007). Adapun tahapan-tahapan pemodelan BMA-normal unimodal adalah sebagai berikut:

**a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian**

Implementasi semua langkah dari Algoritma 2.1 dan Langkah 1 dari Algoritma 5.3 pada data *Chickpea* untuk tujuh ID gen, yaitu LS0035, LS0612, LS0759, U278, U017, LS0132, dan LS0341 disajikan secara lengkap di Lampiran 4. Berdasarkan hasil evaluasi rata-rata dari 5.000 pembangkitan parameter dan 1.000 model yang telah dibentuk dapat diketahui bahwa model tersebut sudah sangat baik dengan terpenuhinya sifat *independence*, *irreducible*, *aperiodic* dan *recurrent* sebagaimana yang telah ditunjukkan pada Gambar 5.14 sampai dengan Gambar 5.17 serta gambar-gambar pada Lampiran 4. Oleh karena itu telah diperoleh model BUM-normal terbaik untuk tujuh ID gen data *Chickpea*.

**b. Seleksi Model dengan Metode Occam's Window**

Langkah 3 dari Algoritma 5.3 akan melakukan seleksi model untuk 1.000 model BUM-normal data *Chickpea* yang telah dibangun. Seleksi untuk 1.000 model BMA-normal unimodal terbaik dari data *Chickpea* dengan metode seleksi Occam's Window dilakukan dengan mengimplementasikan Algoritma 2.3. Hasil yang diperoleh disajikan secara lengkap dalam Tabel 5.10.

Berdasarkan Tabel 5.10 dapat diketahui bahwa pada pemodelan BMA-normal unimodal untuk data *Chickpea* harus dibangun dengan melibatkan 981 model yang terseleksi untuk masing-masing tujuh ID gen yang diamati. Model BMA-normal unimodal yang terbangun cukup kompleks, tetapi dengan kemajuan teknologi komputasi yang maju dan berkembang seperti saat ini, maka kompleksitas model bukan menjadi masalah dalam estimasi modelnya.

**c. Validasi Model BMA-Normal Unimodal**

Untuk memvalidasi model BMA-normal unimodal data *Chickpea* yang telah terbentuk digunakan *goodness of fit* KS. Hasil selengkapnya untuk tujuh ID

gen, yaitu LS0035, LS0612, LS0759, U278, U017, LS0132, dan LS0341 disajikan dalam Tabel 5.11.

Berdasarkan Tabel 5.11 dapat diketahui bahwa model BMA-normal unimodal mampu dengan baik memodelkan tujuh ID gen yang diamati. Hal ini dapat diketahui dari nilai KS yang secara umum memiliki nilai yang cukup kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$ . Untuk model OLS sebagaimana yang telah dijabarkan dalam sub bab 2.2, memiliki nilai KS yang secara umum lebih kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$  dibandingkan dengan nilai KS dari model BMA-normal unimodal.

#### **d. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen**

Pembentukan HPD dengan *credible interval* 95% dilakukan untuk menentukan nilai-nilai batas *credible interval* dari model yang dibangun, sehingga dapat diketahui pengelompokan tujuh ID gen data *Chickpea* ke dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% diperoleh dengan mengimplementasikan Algoritma 5.1. Hasil selengkapnya disajikan dalam Tabel 5.12. Berdasarkan Tabel 5.12 dapat diketahui bahwa ada tiga ID gen yang masuk group *Up-regulated*, yaitu dua ID gen dari fungsi ketahanan, ID gen: LS0612 dan LS0759 serta satu ID gen dari fungsi energi, ID gen: LS0341. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai negatif (-). Sedangkan empat ID gen yang lain masuk group *Regulated*, yaitu tiga ID gen dengan fungsi ketahanan: LS0035, U278, dan U017 serta satu ID gen dengan fungsi energi, ID gen: LS0132. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Sementara tidak satupun ditemukan ID gen yang masuk dalam kategori *Down-regulated*. Hasil yang diperoleh dengan pemodelan BMA-normal unimodal teridentifikasi tiga ID Gen: LS0612, LS0759, dan LS0341 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia.

**Tabel 5.10.** Hasil Seleksi Model dengan Metode Occam's Window untuk Data *Chickpea*

No	ID Gen	Hasil Seleksi 1.000 Model BUM-Normal dengan Metode Occam's Window dalam Pemodelan BMA-Normal Unimodal
Fungsi Ketahanan:		
1	LS0035	981
2	LS0612	981
3	LS0759	981
4	U278	982
5	U017	981
Fungsi Energi:		
6	LS0132	980
7	LS0341	982
Rata-Rata		981,14

**Tabel 5.11.** *Goodness of Fits* KS Model BMA-Normal Unimodal dan Model OLS untuk Data *Chickpea*

No	ID Gen	Kolmogorov-Smirnov (KS)	
		Model OLS	BMA-Normal Unimodal
Fungsi Ketahanan:			
1	LS0035	0,2098	0,2982
2	LS0612	0,3021	0,4543
3	LS0759	0,2937	0,4703
4	U278	0,2528	0,3017
5	U017	0,2711	0,3892
Fungsi Energi:			
6	LS0132	0,2068	0,3805
7	LS0341	0,2087	0,3752

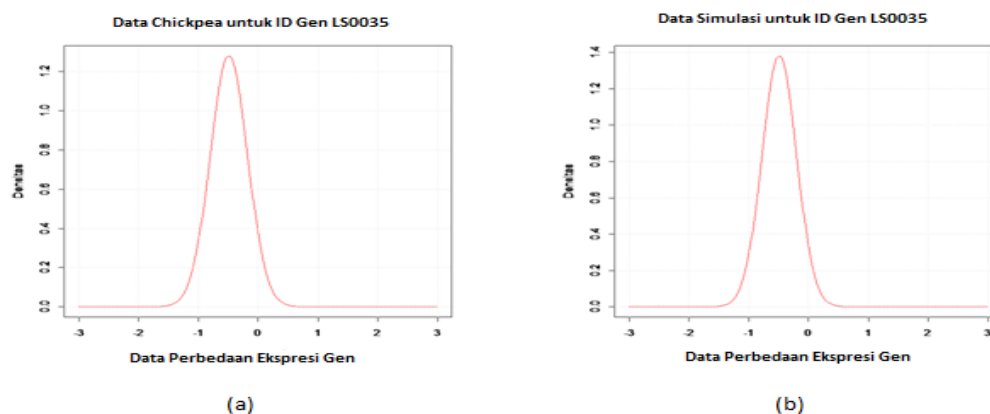
$$KS_{tabel} = 0,708$$

**Tabel 5.12.** HPD dengan *Credible Interval* 95% Model BMA-Normal Unimodal untuk Data *Chickpea*

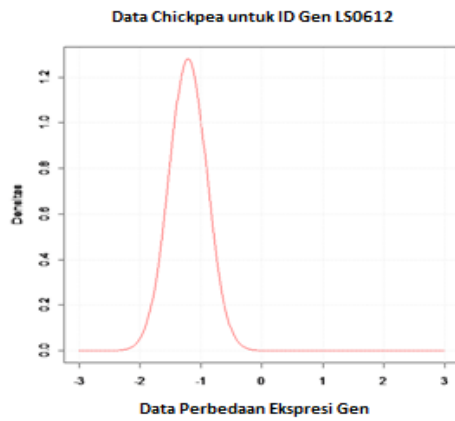
No	ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BMA-Normal Unimodal untuk Data <i>Chickpea</i>		Hasil Pengelompokan
		Batas Bawah	Batas Atas	
Fungsi Ketahanan:				
1	LS0035	-1,1014	0,1221	<i>Regulated</i>
2	LS0612	-1,8242	-0,6021	<i>Up-regulated</i>
3	LS0759	-1,3810	-0,1594	<i>Up-regulated</i>
4	U278	-1,1310	0,0920	<i>Regulated</i>
5	U017	-0,8736	0,3495	<i>Regulated</i>
Fungsi Energi:				
6	LS0132	-1,1333	0,0890	<i>Regulated</i>
7	LS0341	-1,7268	-0,5052	<i>Up-regulated</i>

#### e. Bentuk Distribusi Model BMA-Normal Unimodal

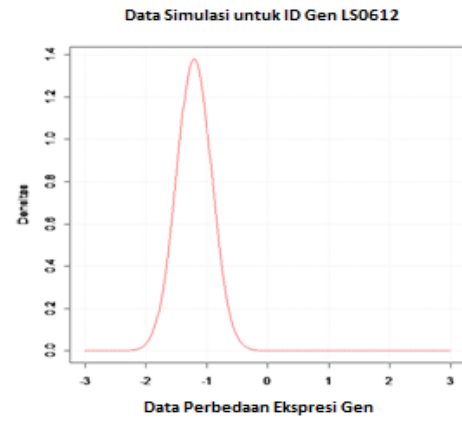
Bentuk distribusi model BMA-normal unimodal data *Chickpea* disajikan bersamaan dengan model BMA-normal unimodal untuk data simulasi karena data simulasi dibangun berdasarkan pola data *Chickpea*. Pemodelan BMA-normal unimodal yang telah dibangun untuk tujuh ID gen memiliki bentuk distribusi normal unimodal yang sama tetapi memiliki karakteristik parameter yang berbeda-beda di setiap ID gen. Selain itu, distribusi data simulasi mengikuti distribusi data *Chickpea*. Bentuk distribusi tujuh ID gen disajikan secara lengkap dalam Gambar 5.18 sampai dengan Gambar 5.24.



**Gambar 5.18.** Bentuk Distribusi BMA-Normal Unimodal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0035

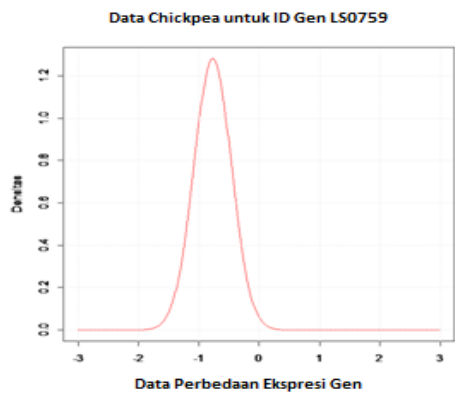


(a)

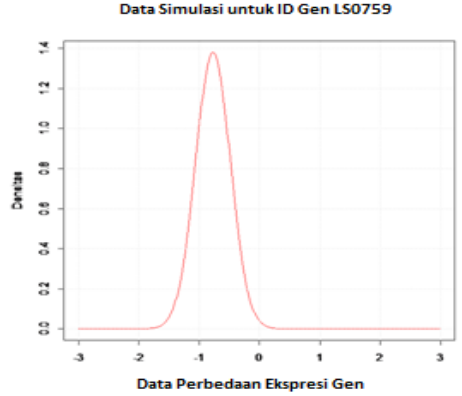


(b)

**Gambar 5.19.** Bentuk Distribusi BMA-Normal Unimodal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0612

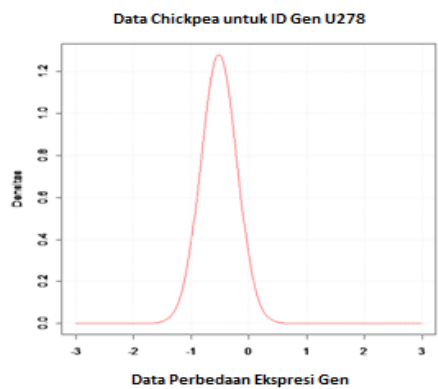


(a)

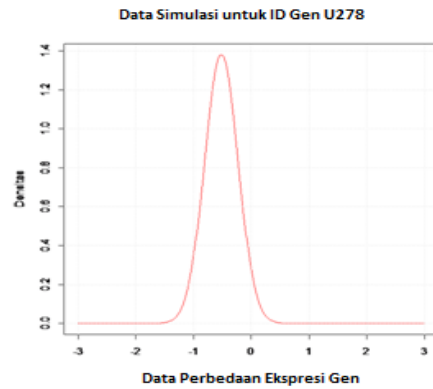


(b)

**Gambar 5.20.** Bentuk Distribusi BMA-Normal Unimodal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0759

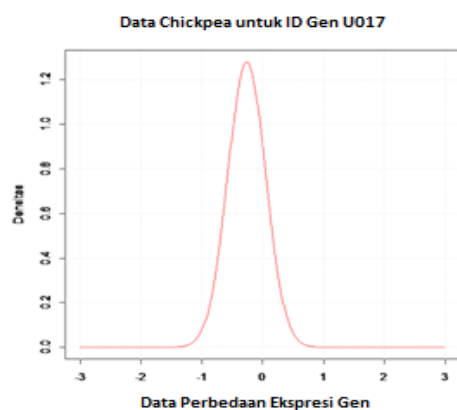


(a)

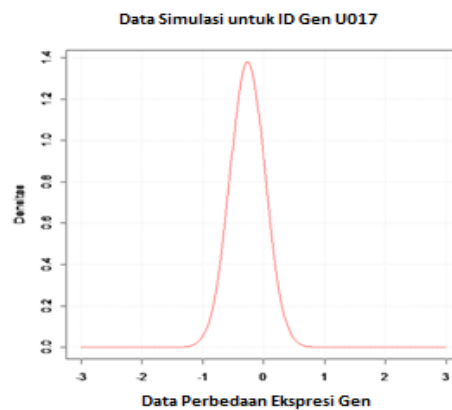


(b)

**Gambar 5.21.** Bentuk Distribusi BMA-Normal Unimodal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U278

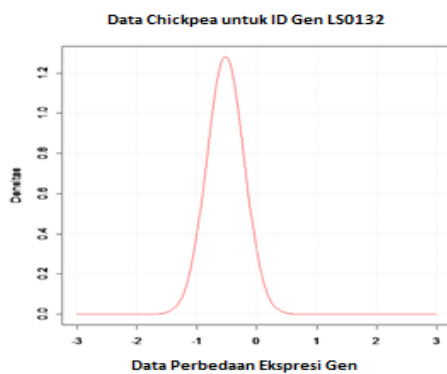


(a)

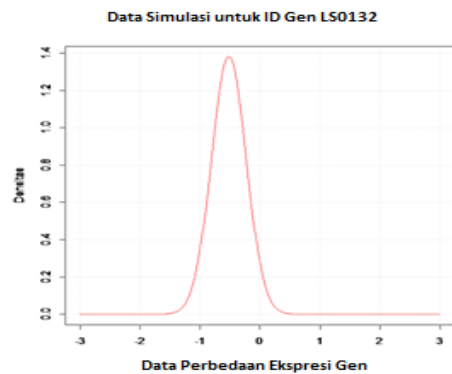


(b)

**Gambar 5.22.** Bentuk Distribusi BMA-Normal Unimodal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U017

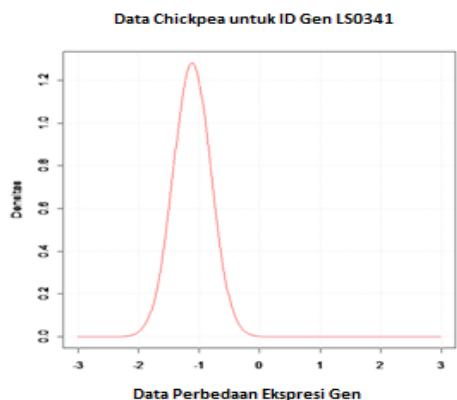


(a)

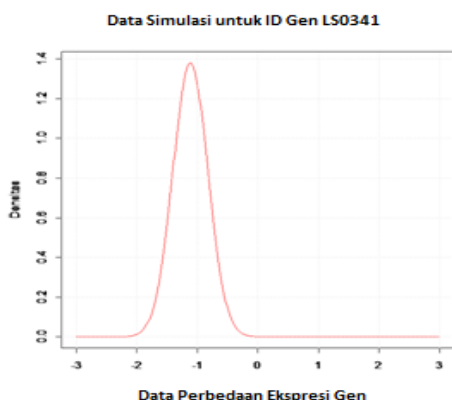


(b)

**Gambar 5.23.** Bentuk Distribusi BMA-Normal Unimodal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0132



(a)



(b)

**Gambar 5.24.** Bentuk Distribusi BMA-Normal Unimodal: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0341

#### 5.4. Perbandingan Hasil Pemodelan BUM-Normal dan BMA-Normal Unimodal

Hasil pemodelan BUM-normal dan BMA-normal unimodal yang telah dibangun akan dibandingkan hasil kebaikan modelnya berdasarkan nilai *goodness of fit* KS dan hasil pengelompokan ID gen. Hasil pengelompokan ID gen asarkan nilai batas HPD dengan *credible interval* 95% yang telah diperoleh sehingga akan diketahui model terbaik di antara model BUM-normal dan BMA-normal unimodal serta hasil pengelompokannya. Tabel 5.13 menyajikan nilai *goodness of fit* KS model BUM-normal dan BMA-normal unimodal untuk data simulasi dan data *Chickpea* beserta hasil pengelompokan ID gen.

Berdasarkan Tabel 5.13 dapat ditunjukkan bahwa model BUM-normal adalah model yang lebih baik dibandingkan dengan model BMA-normal unimodal karena memiliki nilai KS yang secara umum lebih kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$  dibandingkan dengan nilai KS model BMA-normal, baik untuk data simulasi dan maupun data *Chickpea*. Hasil pengelompokan ID gen berdasarkan nilai batas HPD dengan *credible interval* 95% yang telah diperoleh menunjukkan bahwa untuk data simulasi memiliki hasil pengelompokan yang sama antara model BUM-normal dan model BMA-normal unimodal, tetapi untuk data *Chickpea* ada sedikit perbedaan hasil pengelompokan antara model BUM-normal dan model BMA-normal unimodal. Hal ini terjadi pada satu ID gen LS0035 yang teridentifikasi masuk kelompok *Up-regulated* pada model BUM-normal, tetapi pada model BMA-normal unimodal teridentifikasi masuk dalam kelompok *Regulated*. Untuk ID gen yang lain memiliki hasil yang sama antara model BUM-normal dan model BMA-normal unimodal.

Apabila dibandingkan dengan model OLS, maka model OLS secara umum memiliki nilai KS yang paling kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$  baik untuk data simulasi maupun data *Chickpea* dibandingkan dengan model BUM-normal dan BMA-normal unimodal.



**Tabel 5.13.** *Goodness of Fit* KS Model BUM-Normal dan BMA-Normal Unimodal untuk Data Simulasi dan Data *Chickpea* serta Hasil Pengelompokan ID Gen

No	ID Gen	Goodness of Fit Kolmogorov-Smirnov (KS) dan Hasil Pengelompokan ID Gen									
		Model untuk Data Simulasi					Model untuk Data <i>Chickpea</i>				
		OLS	BUM-Normal		BMA-Normal Unimodal		OLS	BUM-Normal		BMA-Normal Unimodal	
Fungsi Ketahanan:	Nilai KS		Hasil Pengelompokan	Nilai KS	Hasil Pengelompokan	Nilai KS		Hasil Pengelompokan	Nilai KS	Hasil Pengelompokan	
1	LS0035	<b>0,2710</b>	0,3589*	<i>Regulated</i>	0,3672	<i>Regulated</i>	<b>0,2098</b>	0,2878*	<i>Up-regulated</i>	0,2982	<i>Regulated</i>
2	LS0612	<b>0,2820</b>	0,4768	<i>Up-regulated</i>	0,4754*	<i>Up-regulated</i>	<b>0,3021</b>	0,4463*	<i>Up-regulated</i>	0,4543	<i>Up-regulated</i>
3	LS0759	<b>0,2841</b>	0,4810	<i>Up-regulated</i>	0,4795*	<i>Up-regulated</i>	<b>0,2937</b>	0,4723	<i>Up-regulated</i>	0,4703*	<i>Up-regulated</i>
4	U278	<b>0,2810</b>	0,3540*	<i>Regulated</i>	0,3613	<i>Regulated</i>	<b>0,2528</b>	0,2914*	<i>Regulated</i>	0,3017	<i>Regulated</i>
5	U017	<b>0,2803</b>	0,4332*	<i>Regulated</i>	0,4362	<i>Regulated</i>	<b>0,2711</b>	0,3845*	<i>Regulated</i>	0,3892	<i>Regulated</i>
Fungsi Energi:											
6	LS0132	<b>0,2819</b>	0,4436*	<i>Regulated</i>	0,4464	<i>Regulated</i>	<b>0,2068</b>	0,3711*	<i>Regulated</i>	0,3805	<i>Regulated</i>
7	LS0341	<b>0,2874</b>	0,4307*	<i>Up-regulated</i>	0,4344	<i>Up-regulated</i>	<b>0,2087</b>	0,3706*	<i>Up-regulated</i>	0,3752	<i>Up-regulated</i>

$$KS_{tabel} = 0,708$$

## BAB VI

### IDENTIFIKASI PERBEDAAN EKSPRESI GEN TANAMAN *CHICKPEA* DI INDONESIA DENGAN PEMODELAN BAYESIAN *MIXTURE* MODEL

Dalam bab ini akan disajikan beberapa algoritma yang terkait dengan identifikasi perbedaan ekspresi gen tanaman *Chickpea* di Indonesia melalui pemodelan Bayesian *Mixture* Normal (BMM-normal) dan implementasi beberapa algoritma yang telah dibangun dilakukan pada data simulasi dan data *Chickpea*.

#### 6.1. Pemodelan dan Algoritma untuk Bayesian *Mixture* Model (BMM) Normal

Pemodelan Bayesian *mixture* merupakan sebuah pemodelan yang menggunakan pendekatan Bayesian terhadap suatu data yang polanya memiliki modus lebih dari satu (multimodal). Data yang memiliki distribusi multimodal akan memiliki susunan sub-sub populasi atau group-group populasi dimana setiap sub populasi merupakan komponen penyusun dari model *mixture* yang memiliki proporsi berbeda-beda. Dasar perhitungan dengan konsep Bayesian adalah menentukan distribusi posterior parameter model berdasarkan perkalian dari fungsi *likelihood* dan distribusi prior. Pemodelan BMM-normal dibentuk dengan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* sebagaimana yang telah dijabarkan pada Algoritma 2.1.

Pada Algoritma 2.1, parameter  $\theta$  yang dimaksud adalah  $\lambda[i]$ ,  $\tau[i]$ , dan  $w[i]$ , dengan  $i = 1, 2, \dots, k$ , dimana  $k$  merupakan jumlah komponen dalam distribusi *mixture* normal. Berdasarkan Algoritma 2.1 akan diperoleh penduga-penduga rata-rata bagi  $\lambda[i]$ ,  $\tau[i]$ , dan  $w[i]$  untuk  $\forall i$  dalam pemodelan BMM-normal dan melakukan evaluasi model untuk mengetahui pemenuhan asumsi model dengan pendekatan Bayesian.

Pemodelan BMM-normal dibangun untuk data yang berdistribusi *mixture* normal, sehingga untuk membentuk model *mixture* normal ini dibutuhkan

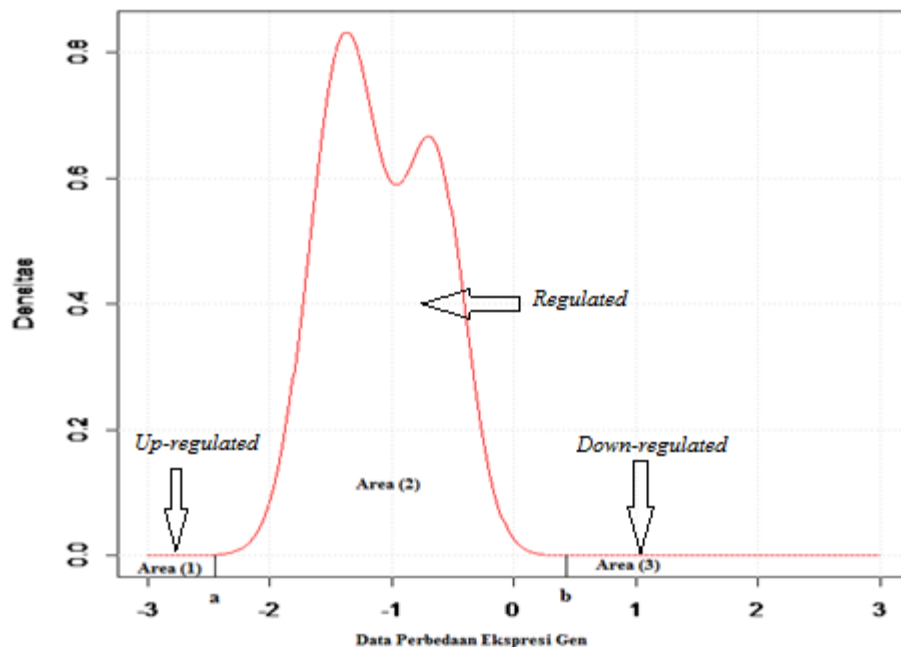
informasi fungsi *likelihood* dari data sebagaimana yang telah dijabarkan secara rinci pada sub bab 2.4. Selain itu juga dibutuhkan informasi tentang distribusi prior yang paling tepat untuk masing-masing parameter dalam model *mixture* normal. Penentuan distribusi prior ini menggunakan kriteria sebagaimana yang telah dibahas secara detail dalam sub bab 2.4.2. Langkah selanjutnya adalah menentukan distribusi posterior parameter model *mixture* normal yang berupa distribusi gabungan dari semua variabel pada model *mixture* normal. Penjabaran tentang hal ini, termasuk distribusi bersyarat penuh dari setiap parameter model BMM-normal, juga telah dijabarkan secara rinci pada sub bab 2.4.

Untuk lebih memudahkan estimasi parameter model BMM-normal, maka struktur hirarki dari distribusi posterior model BMM-normal dapat digambarkan dalam bentuk *Directed Acyclic Graph* (DAG) sebagaimana yang ditunjukkan pada Gambar 2.1. Melalui bantuan program WinBUGS, estimasi parameter model BMM-normal dapat diketahui.

Hasil model BMM-normal yang diperoleh kemudian dievaluasi pemenuhan asumsinya dalam pemodelan Bayesian, yaitu apakah telah memenuhi kriteria *independence* untuk kebebasan antaraestimasi antariterasi dalam sampel MCMC, *irreducible*, *aperiodic* dan *recurrent* untuk konvergensi estimasi parameter. Hal ini dapat dilihat dari plot fungsi autokorelasi yang berciri lag-0 nyata berbeda dengan nol dan lag-lag selain lag-0 tidak berbeda nyata dengan nol (sifat *independence* antara estimasi antar iterasi telah terpenuhi) dan plot distribusi Kernel, *history* dan *quantiles* dari masing-masing parameter yang berciri *irreducible*, *aperiodic* dan *recurrent* yang menunjukkan konvergensi estimasi parameter telah terpenuhi.

Apabila model BMM-normal terbaik telah diperoleh, maka langkah selanjutnya memvalidasi model BMM-normal dengan kriteria *goodness of fit* KS. Dalam hal ini digunakan konsep jarak terbesar antara densitas kumulatif empiris dan densitas kumulatif model BMM-normal, dimana semakin kecil nilai KS yang diperoleh dan masih dalam jaminan lebih kecil dari  $KS_{tabel}$ , maka model yang dibangun makin baik. Konsep *goodness of fit* KS ini telah dijelaskan secara lengkap dalam sub bab 2.8.

Langkah terakhir dari pemodelan BMM-normal adalah membentuk HPD dengan *credible interval* 95% untuk menentukan nilai-nilai batas *credible interval*, sehingga data dapat dikelompokkan ke dalam tiga kelompok, yaitu *Down-regulated*, *Regulated*, dan *Up-regulated* menggunakan Algoritma 2.3 yang diterapkan pada fungsi densitas  $f(\mathbf{x}|\boldsymbol{\theta})$ , *mixture* normal. Dalam pembentukan HPD ini dibutuhkan informasi model BMM-normal terbaik dari data sehingga nilai-nilai batas tersebut dapat ditetapkan. Berikut ini disajikan contoh HPD dengan *credible interval* 95% dari model BMM-normal (Gambar 6.1). Area (1) merupakan area untuk kelompok data *Up-regulated*, Area (2) untuk kelompok data *Regulated*, dan Area (3) untuk kelompok data *Down-regulated*.



**Gambar 6.1.** Ilustrasi HPD dengan *Credible Interval* 95% untuk Model BMM-normal

Untuk memudahkan proses pembentukan model BMM-normal maka dibangun Algoritma 6.1 yang merupakan Algoritma BMM-normal.

#### **Algoritma 6.1.** Algoritma BMM-Normal

Tahapan-tahapan dari Algoritma BMM-normal melalui langkah-langkah berikut:

- Langkah 1. Susun model BMM-normal rata-rata dari 1.000 model yang dibangun menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* dengan jumlah iterasi parameter yang diduga dalam model sebanyak 5.000 kali
- Langkah 2. Evaluasi pemenuhan asumsi pemodelan Bayesian (*independence, irreducible, aperiodic, dan recurrent*) model BMM-normal rata-rata yang terbaik dari hasil Langkah 1 melalui kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diduga dalam model.
- Langkah 3. Hitung nilai *goodness of fit* KS untuk validasi model BMM-normal sebagaimana yang telah dijabarkan pada sub bab 2.8
- Langkah 4. Tentukan nilai batas bawah (namakana) dan batas atas (namakan b) melalui HPD dengan *credible interval* 95% sebagaimana yang telah dijabarkan pada sub bab 2.9.
- Langkah 5. Tetapkan kelompok data ke dalam tiga kelompok, yaitu *Down-regulated, Regulated, dan Up-regulated*.

## 6.2. Implementasi pada Data Simulasi

Implementasi Algoritma 6.1 dilakukan pada data simulasi untuk membentuk model BMM-normal, sehingga dapat diketahui sejauh mana kebaikan model BMM-normal yang telah dibangun untuk data simulasi. Tahapan-tahapan pemodelan diuraikan sebagai berikut:

### a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian

Berdasarkan data simulasi yang telah dibangun dengan ciri data *Chickpea* sebagaimana yang telah dijabarkan pada sub bab 4.4, teridentifikasi satu ID gen dari fungsi ketahanan memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen: U070 dan sembilan ID gen memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, dan U273. Sedangkan ID gen dari fungsi energi, ada dua ID gen yang memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen: U001

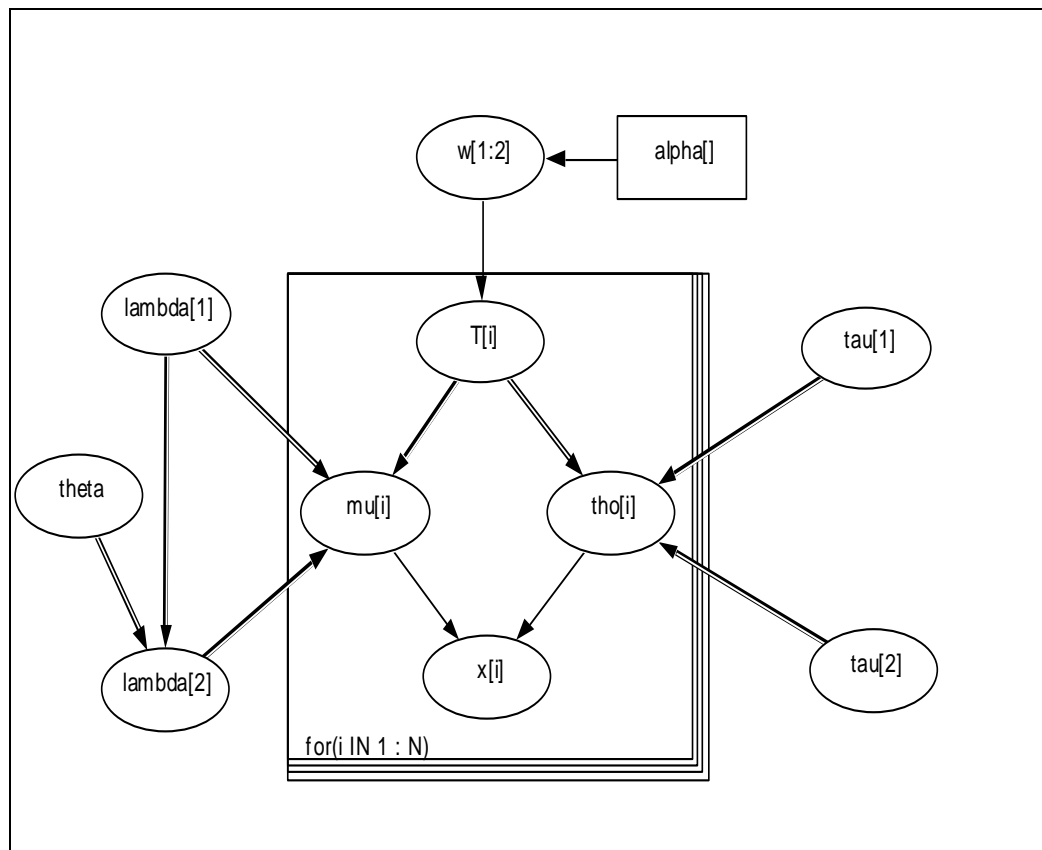
dan CA0810, dan tujuh ID gen memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: U325, U385, U447, CA0426, CA0534, CA0702, dan CA1120. Dalam hal ini berarti bahwa dari 26 ID gen pada data simulasi, ada tiga ID gen yang memiliki distribusi *mixture* normal tiga komponen dan 16 ID gen yang memiliki distribusi *mixture* normal dua komponen. Hal ini berarti total 19 ID gen yang akan dimodelkan *mixture* normal dengan pendekatan Bayesian (BMM-normal) menggunakan implementasi Algoritma 6.1. Sedangkan tujuh ID gen yang lain, yaitu ID gen: LS0035, LS0612, LS0759, U278, U017, LS0132, dan LS0341 tidak dimodelkan BMM-normal karena memiliki distribusi normal unimodal (bukan *mixture*).

Estimasi parameter pada model BMM-normal dilakukan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) dengan distribusi *mixture* normal dua dan tiga komponen. Implementasi Algoritma 2.1 ini menggunakan program WinBUGS dengan membuat *Syntax BUGS* atau *Graphical Model* (Doodle BUGS). Berikut ini disajikan *Graphical Model* dan *Syntax BUGS* bagi model BMM-normal untuk distribusi *mixture* normal dua dan tiga komponen (Gambar 6.2 dan Gambar 6.3 serta *Syntax* 6.1 dan *Syntax* 6.2) bagi 19 ID gen benar telah teridentifikasi memiliki distribusi *mixture* normal.

Hasil *running* program BUGS, diberikan contoh ID gen LS0024 untuk model *mixture* normal dua komponen (Tabel 6.1 dan Gambar 6.4 sampai dengan Gambar 6.7) dan ID gen U070 untuk model *mixture* normal tiga komponen (Tabel 6.2 dan Gambar 6.8 sampai dengan Gambar 6.11). Hasil yang disajikan merupakan rata-rata data simulasi untuk pembangkitan data sebanyak 1.000 kali, dimana iterasi setiap parameter dilakukan sebanyak 5.000 kali. Hasil secara lengkap untuk 19 ID Gen yang lain, yaitu LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 dengan fungsi ketahanan, serta ID gen U325, U385, U447, CA0426, CA0534, CA0702, CA1120, U001, dan CA0810 dengan fungsi energi, dapat dilihat pada Lampiran 5.

Berdasarkan hasil evaluasi estimasi parameter untuk model BMM-normal dua dan tiga komponen dapat diketahui bahwa untuk 19 ID gen dari data simulasi yang diamati sudah sangat baik dan sudah memenuhi asumsi pemodelan pendekatan Bayesian, yaitu *independence*, *irreducible*, *aperiodic* dan *recurrent*

dalam sampel MCMC untuk mencapai konvergensi estimasinya. Hal ini dapat diketahui dari plot fungsi autokorelasi semua parameter yang diestimasi mempunyai nilai nyata berbeda dengan nol untuk lag-0 dan lag-lag selain lag-0 tidak nyata berbeda dengan nol (Gambar 6.4 dan Gambar 6.8). Demikian juga dengan plot distribusi Kernel, *history*, dan *quantiles* dari parameter yang diestimasi memiliki ciri-ciri yang sesuai dengan model yang diharapkan, yaitu estimasi parameter telah mencapai konvergen dengan memenuhi sifat *irreducible*, *aperiodic* dan *recurrent* (Gambar 6.5 sampai dengan Gambar 6.7 dan Gambar 6.9 sampai dengan Gambar 6.11). Oleh karena itu, model BMM-normal dua dan tiga komponen terbaik untuk 19 ID gen data simulasi telah diketahui.



**Gambar 6.2.** *Graphical Model (Doodle BUGS)* untuk Model BMM-Normal Dua Komponen

### Syntax 6.1. Syntax BUGS untuk Model BMM-Normal Dua Komponen

```

model;
{
  w[1:2] ~ ddirch(alpha[])
  for( i in 1 : N ) {
    T[i] ~ dcat(w[1:2])
  }
  for( i in 1 : N ) {
    mu[i] <- lambda[T[i]]
  }
  for( i in 1 : N ) {
    x[i] ~ dnorm(mu[i],tho[i])
  }
  lambda[1] ~ dnorm( 0.1,0.05)
  lambda[2] <- lambda[1] + theta
  theta ~ dnorm( 0.0,1)
  for( i in 1 : N ) {
    tho[i] <- tau[T[i]]
  }
  tau[2] ~ dgamma(1,1)
  tau[1] ~ dgamma(1,1)
}

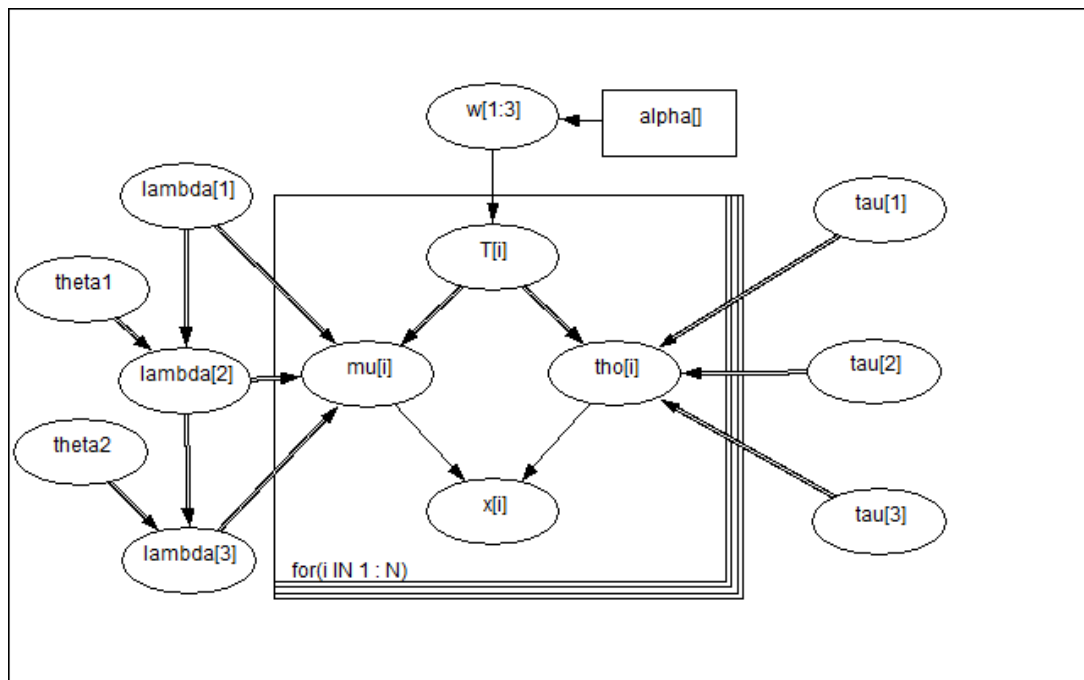
#Update tool thin 10 untuk semua ID Gen
#Analisis untuk 16 ID Gen, model BMM-Normal 2 Komponen

#1.Data Simulasi: LS0024
DATA
list(x=c(-0.6241,-0.4632,-0.088),N=3,alpha=c(1,1),T=c(1,1,2))

#1.Data Harijati: LS0024
DATA
list(x=c(-0.6774,-0.4635,-0.0345),N=3,alpha=c(1,1),T=c(1,1,2))

INITIAL
list(lambda=c(0.0,NA),theta=0.1,w=c(0.6,0.4)))

```



**Gambar 6.3.** Graphical Model (Doodle BUGS) untuk Model BMM-Normal Tiga Komponen



### Syntax 6.2. Syntax BUGS untuk Model BMM-Normal Tiga Komponen

```

model;
{
  w[1:3] ~ ddirch(alpha[])
  for( i in 1 : N ) {
    T[i] ~ dcat(w[1:3])
  }
  for( i in 1 : N ) {
    mu[i] <- lambda[T[i]]
  }
  for( i in 1 : N ) {
    x[i] ~ dnorm(mu[i],tho[i])
  }
  lambda[1] ~ dnorm( 5.0,0.574)
  lambda[2] <- lambda[1] + theta1
  lambda[3] <- lambda[2] + theta2
  theta1 ~ dunif(0,1000)
  theta2 ~ dunif(0,1000)
  for( i in 1 : N ) {
    tho[i] <- tau[T[i]]
  }
  tau[2] ~ dgamma(15,1)
  tau[1] ~ dgamma(11.98,1)
  tau[3] ~ dgamma(15.01,1)
}

#Update tool thin 10 untuk semua ID Gen
#Analisis untuk 3 ID Gen, model BMM-Normal 3 Komponen
#-----

#1.Data Simulasi: U070
DATA
list(x=c(-1.4864,-0.7061,-0.2483),N=3,alpha=c(1,1,1),T=c(1,2,3))

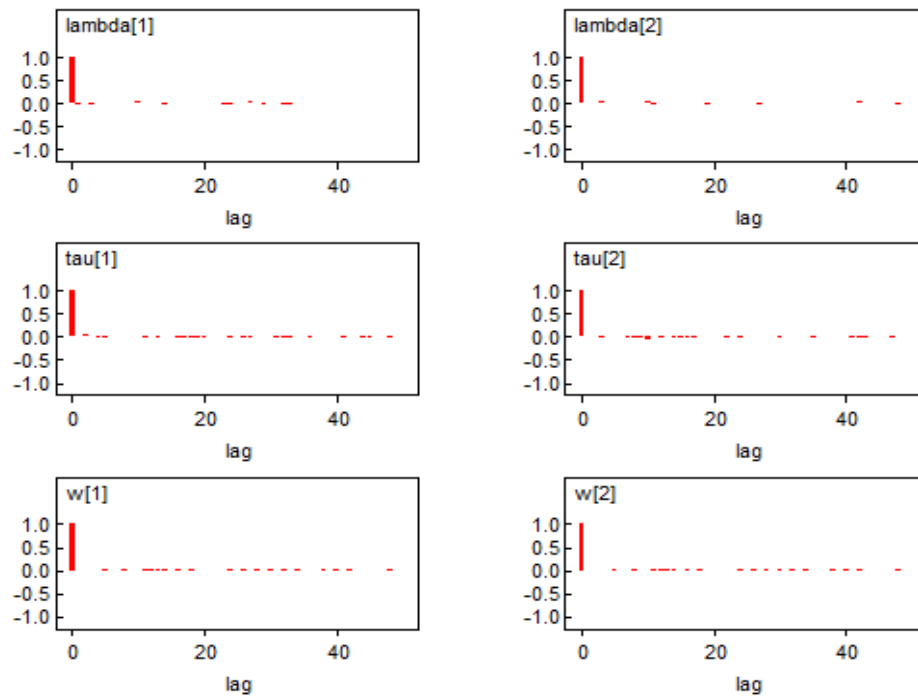
#1.Data Harijati: U070
DATA
list(x=c(-0.7078,-1.5376,-0.1971),N=3,alpha=c(1,1,1),T=c(1,2,3))

INITIAL
list(lambda=c(2.0,NA,NA),theta1=0.1,theta2=0.1,w=c(0.4,0.3,0.3)))

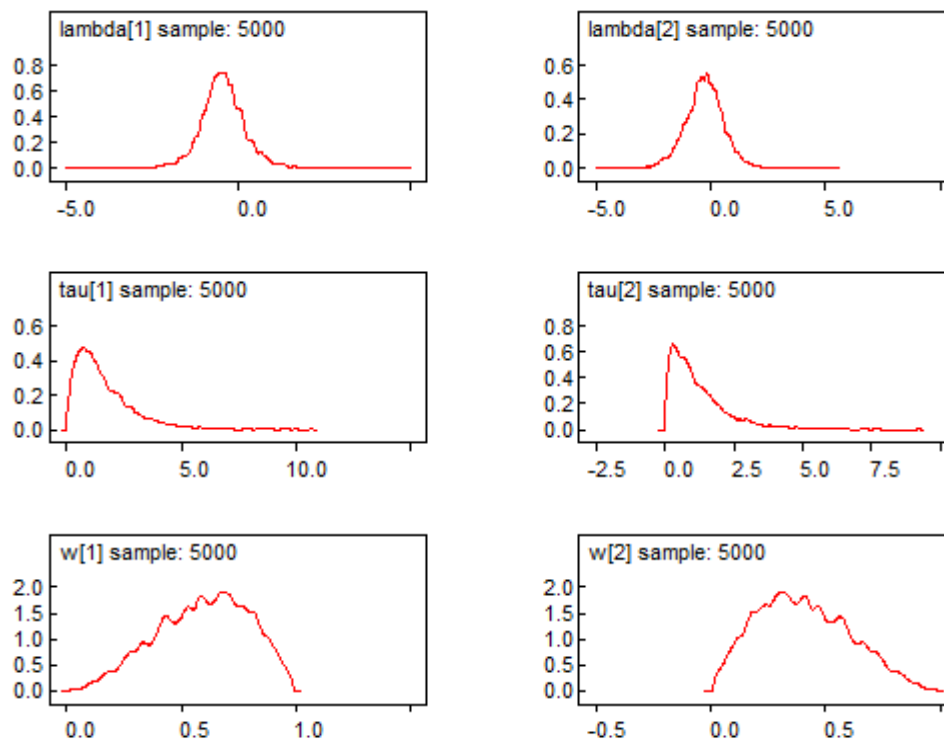
```

**Tabel 6.1.** Estimasi Parameter Rata-Rata Data Simulasi ID Gen LS0024 dengan Model BMM-Normal Dua Komponen

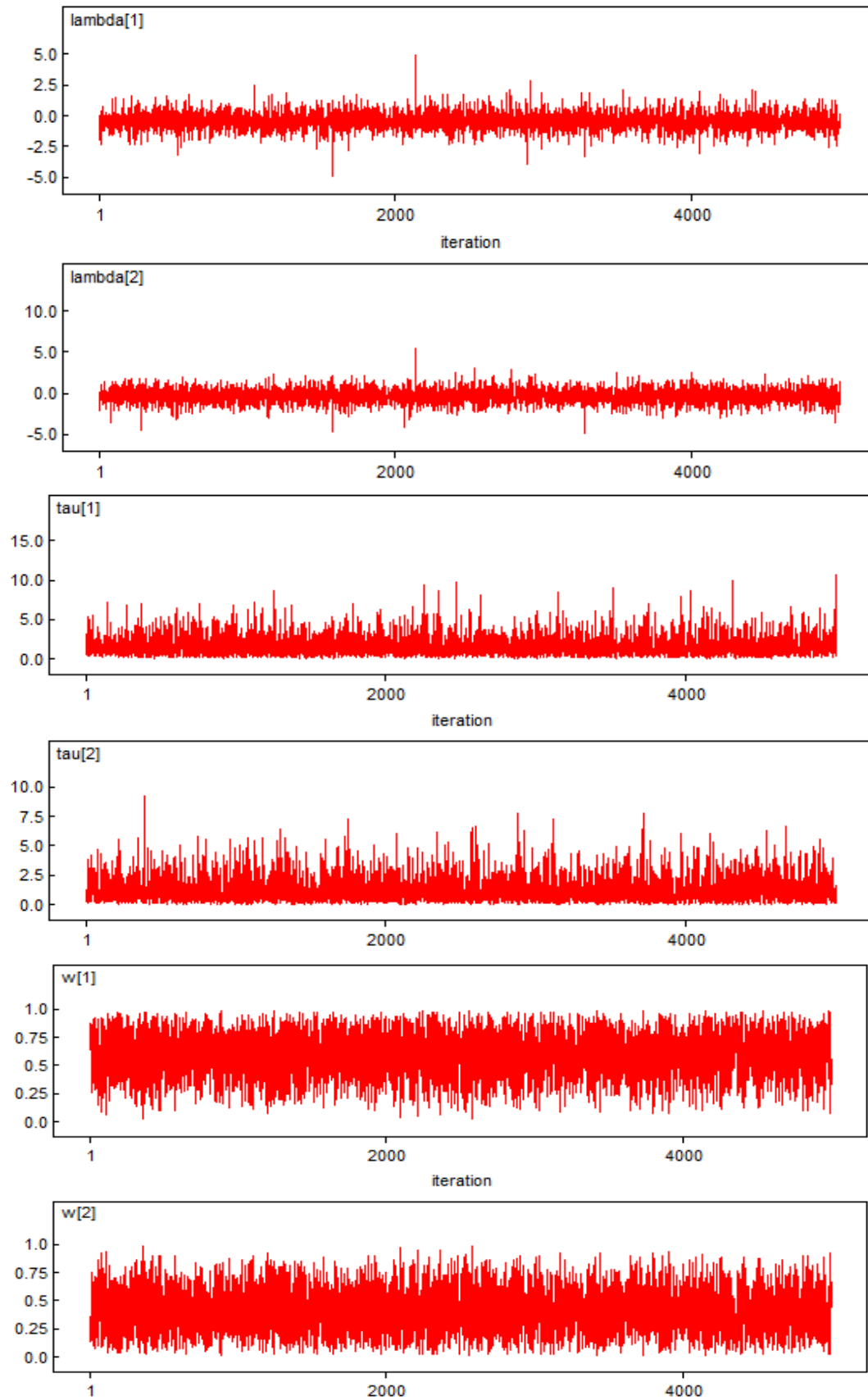
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,4487	0,6434	0,0079	-1,713	-0,4705	0,8975	1	5.000
lambda[2]	-0,2848	0,8308	0,0111	-2,016	-0,2637	1,328	1	5.000
tau[1]	1,585	1,229	0,0178	0,1573	1,271	4,76	1	5.000
tau[2]	1,198	1,04	0,0134	0,0827	0,9056	4,035	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000



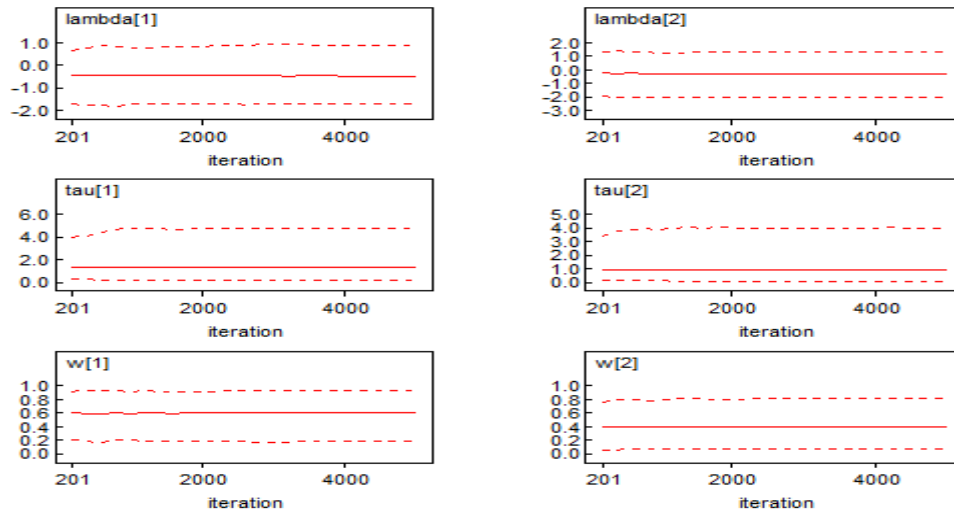
**Gambar 6.4.** Fungsi Autokorelasi setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024



**Gambar 6.5.** Densitas Kernel setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024



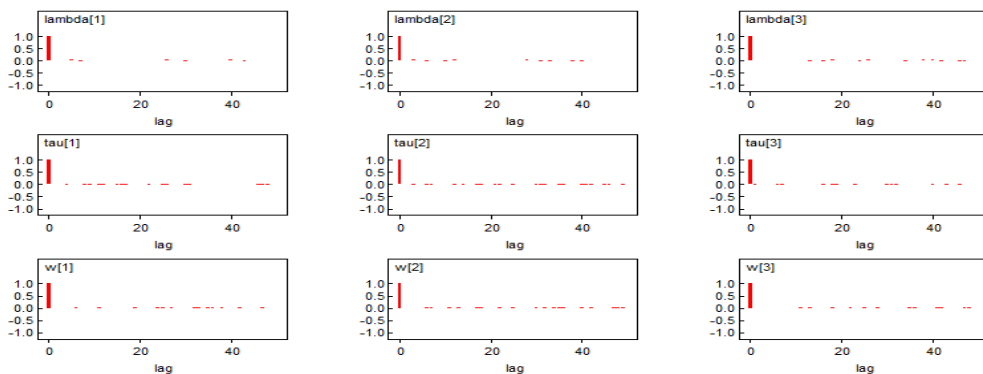
**Gambar 6.6.** *History Data* Bangkitan setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024



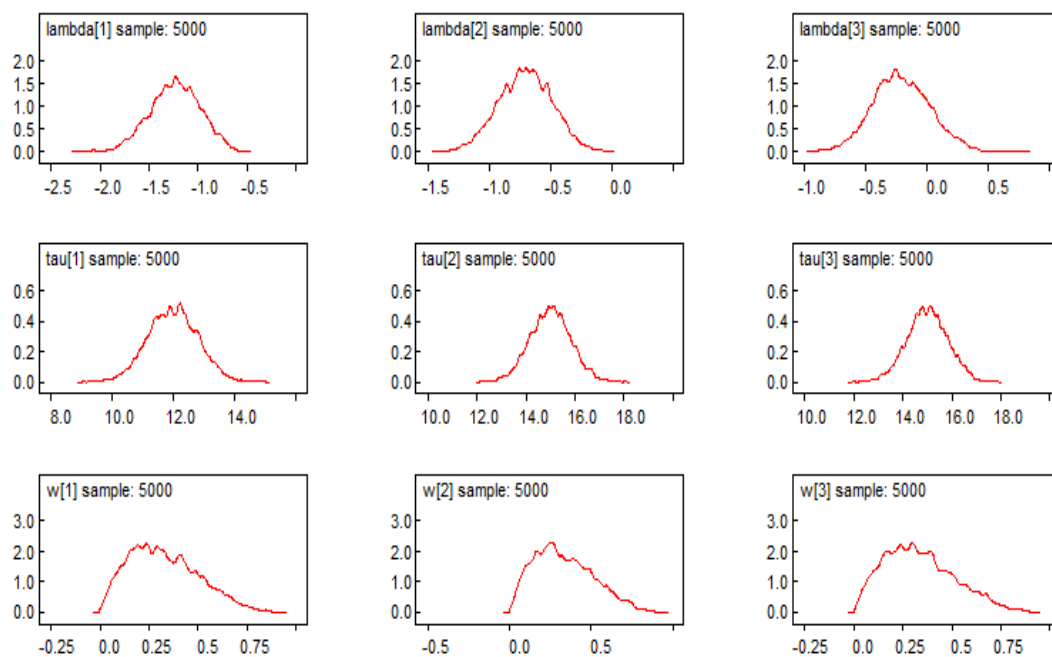
**Gambar 6.7.** *Quantiles* setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024

**Tabel 6.2.** Estimasi Parameter Rata-Rata Data Simulasi ID Gen U070 dengan Model BMM-Normal Tiga Komponen

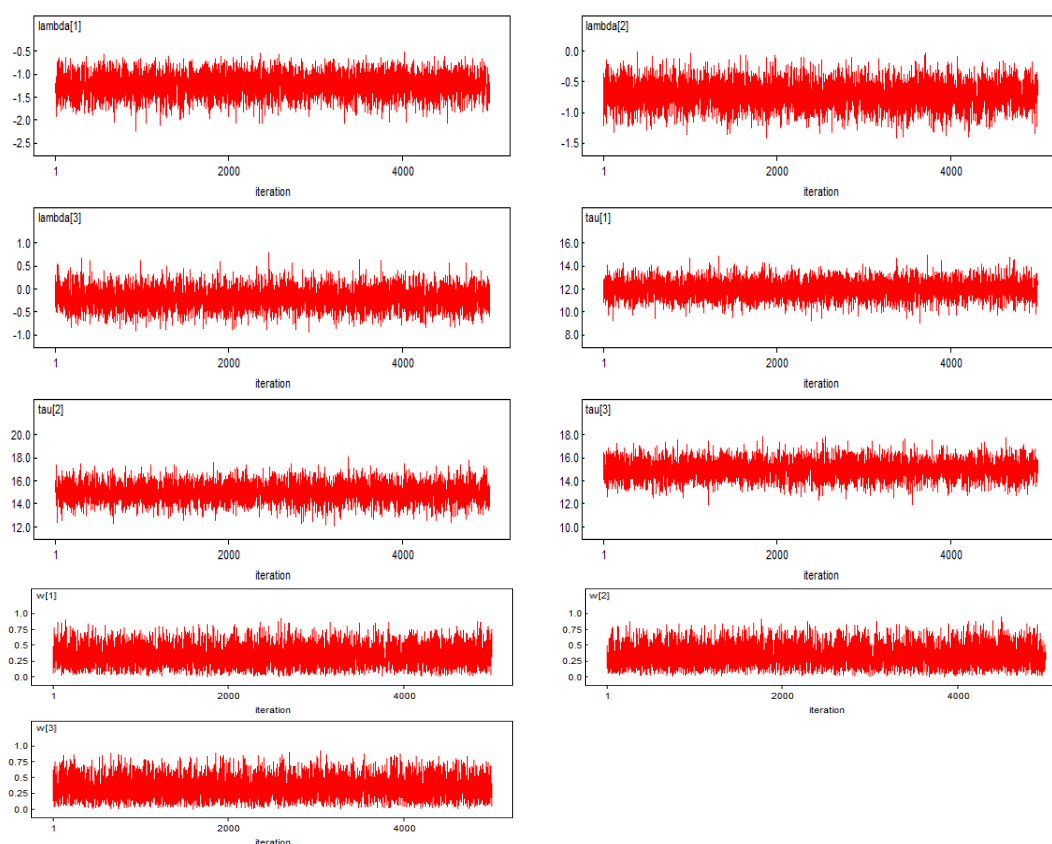
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,227	0,2654	0,0042	-1,766	-1,223	-0,7236	1	5.000
lambda[2]	-0,7054	0,2246	0,0034	-1,143	-0,7062	-0,2705	1	5.000
lambda[3]	-0,2022	0,2394	0,0036	-0,6557	-0,2043	0,2863	1	5.000
tau[1]	11,72	3,43	0,0448	6,055	11,32	19,55	1	5.000
tau[2]	15,05	3,912	0,0551	8,521	14,6	24,11	1	5.000
tau[3]	15,07	3,807	0,0610	8,565	14,71	23,43	1	5.000
w[1]	0,3318	0,1785	0,0023	0,0498	0,3107	0,7137	1	5.000
w[2]	0,3334	0,1823	0,0024	0,0546	0,3112	0,7358	1	5.000
w[3]	0,3348	0,1797	0,0023	0,0535	0,3158	0,7141	1	5.000



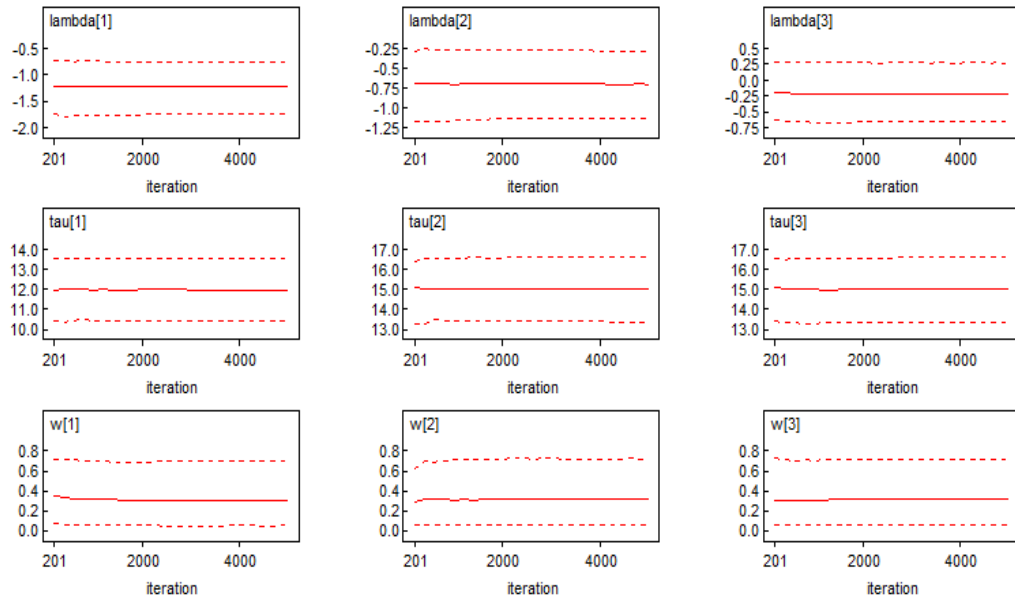
**Gambar 6.8.** Fungsi Autokorelasi setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070



**Gambar 6.9.** Densitas Kernel setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070



**Gambar 6.10.** History Data Bangkitan setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070



**Gambar 6.11.** *Quantiles* setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070

## b. Validasi Model BMM-Normal

Untuk memvalidasi model BMM-normal yang telah terbentuk digunakan *goodness of fit* KS. Hasil selengkapnya untuk 19 ID gen, yaitu LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 sebagai pembawa fungsi ketahanan serta ID gen U325, U385, U447, CA0426, CA0534, CA0702, CA1120, U001, dan CA0810 sebagai pembawa fungsi energi disajikan dalam Tabel 6.3.

Berdasarkan hasil validasi model data simulasi dengan *goodness of fit* KS (Tabel 6.3) dapat diketahui bahwa model BMM-normal dua dan tiga komponen mampu memodelkan 19 ID gen yang diamati dengan baik. Hal ini dapat diketahui dari nilai rata-rata KS yang secara umum memiliki nilai cukup kecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel} = 0,708$ . Untuk model OLS sebagaimana yang telah dijabarkan pada sub bab 2.2, memiliki nilai rata-rata KS yang secara umum lebih kecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel} = 0,708$  dibandingkan dengan nilai rata-rata KS model BMM-normal.

### c. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen

HPD dengan *credible interval* 95% dibentuk untuk menentukan nilai-nilai batas *credible interval*, sehingga dapat diketahui pengelompokan dari 19 ID gen yang diamati ke dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% diperoleh dengan mengimplementasikan Algoritma 2.3. Hasil pembuatan HPD selengkapnya disajikan dalam Tabel 6.4.

Berdasarkan Tabel 6.4 dapat diketahui bahwa ada tujuh ID gen yang masuk ke dalam group *Up-regulated*, empat ID gen dari fungsi ketahanan, yaitu ID gen: LS0081, LS0752, LS0688, dan U273 serta tiga ID gen dari fungsi energi, yaitu ID gen: U325, CA0534, dan CA0702. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai negatif (-). Sedangkan 12 ID gen yang lain masuk group *Regulated*, yaitu enam ID gen dengan fungsi ketahanan: LS0024, LS0616, LS0159, LS0162, LS0185, dan U070 serta enam ID gen dengan fungsi energi, ID gen: U385, U447, CA0426, CA1120, U001, dan CA0810. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Sementara itu, tidak ada satupun ID gen yang masuk ke dalam kategori *Down-regulated*. Hasil yang diperoleh dengan pemodelan BMM-normal dua dan tiga komponen teridentifikasi tujuh ID Gen: LS0081, LS0752, LS0688, U273, U325, CA0534, dan CA0702 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia.

### d. Bentuk Distribusi Model BMM-Normal Dua dan Tiga Komponen

Bentuk distribusi model BMM-normal dua dan tiga komponen dari data *Chickpea* disajikan bersamaan dengan model BMM-normal dua dan tiga komponen untuk data simulasi karena data simulasi dibangun berdasarkan pola data *Chickpea*. Pemodelan BMM-normal dua dan tiga komponen yang telah dibangun untuk 19 ID gen memiliki bentuk distribusi *mixture* normal yang sama tetapi memiliki nilai estimasi parameter yang berbeda-beda di setiap ID gen. Selain itu, distribusi data simulasi mengikuti distribusi data *Chickpea*. Bentuk

distribusi 19 ID gen, disajikan secara lengkap dalam Gambar 6.12 sampai dengan Gambar 6.30.

**Tabel 6.3.** Rata-Rata Nilai *Goodness of Fit* KS Model BMM-Normal Dua dan Tiga Komponen untuk Model OLS Data Simulasi

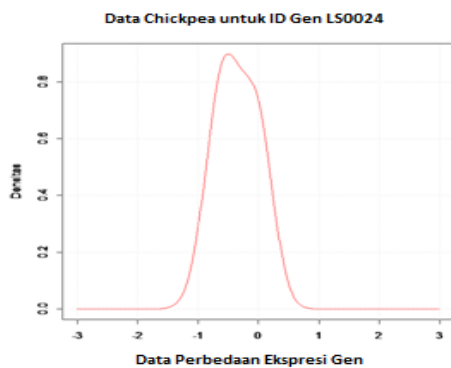
No	ID Gen	Nilai Rata-Rata Kolmogorov-Smirnov (KS) dari 1.000 kali Pembangkitan	
		Model OLS	BMM-Normal
Fungsi Ketahanan:			
1	LS0024	0,2670	0,2785
2	LS0616	0,3479	0,3168
3	LS0159	0,3058	0,2935
4	LS0162	0,2793	0,2791
5	LS0185	0,2877	0,2861
6	LS0081	0,3465	0,3181
7	LS0752	0,3417	0,3163
8	LS0688	0,2637	0,2725
9	U273	0,3477	0,3227
10	U070	0,2344	0,2699
Fungsi Energi:			
11	U325	0,3275	0,3161
12	U385	0,2774	0,2894
13	U447	0,3318	0,3183
14	CA0426	0,2986	0,2769
15	CA0534	0,2896	0,3891
16	CA0702	0,2892	0,3044
17	CA1120	0,3684	0,3451
18	U001	0,2304	0,2692
19	CA0810	0,2328	0,3692

$$KS_{tabel} = 0,708$$

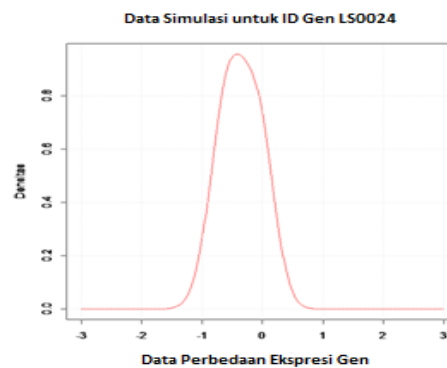


**Tabel 6.4.** HPD dengan *Credible Interval* 95% untuk Model BMM-Normal dengan Dua dan Tiga Komponen Data Simulasi

No	ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BMM-Normal Data Simulasi		Hasil Pengelompokan
		Batas Bawah	Batas Atas	
Fungsi Ketahanan:				
1	LS0024	-1,0667	0,3367	<i>Regulated</i>
2	LS0616	-1,1907	0,0860	<i>Regulated</i>
3	LS0159	-1,4059	0,0604	<i>Regulated</i>
4	LS0162	-1,1218	0,3330	<i>Regulated</i>
5	LS0185	-1,0779	0,3356	<i>Regulated</i>
6	LS0081	-1,3770	-0,0372	<i>Up-regulated</i>
7	LS0752	-1,5493	-0,1917	<i>Up-regulated</i>
8	LS0688	-1,4899	-0,0682	<i>Up-regulated</i>
9	U273	-2,1536	-0,8466	<i>Up-regulated</i>
10	U070	-1,6090	0,1704	<i>Regulated</i>
Fungsi Energi:				
11	U325	-1,8757	-0,2699	<i>Up-regulated</i>
12	U385	-1,1874	0,1664	<i>Regulated</i>
13	U447	-1,1553	0,3132	<i>Regulated</i>
14	CA0426	-1,4357	0,3143	<i>Regulated</i>
15	CA0534	-1,3569	-0,0987	<i>Up-regulated</i>
16	CA0702	-1,4004	-0,0394	<i>Up-regulated</i>
17	CA1120	-1,2127	0,7105	<i>Regulated</i>
18	U001	-1,1626	0,4169	<i>Regulated</i>
19	CA0810	-1,2676	0,0674	<i>Regulated</i>

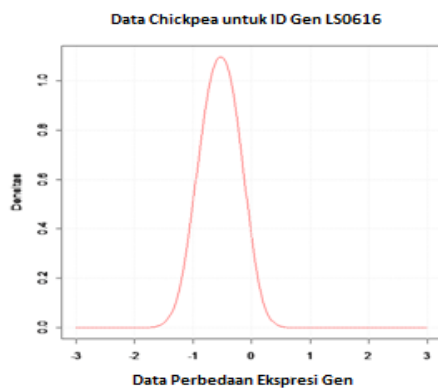


(a)

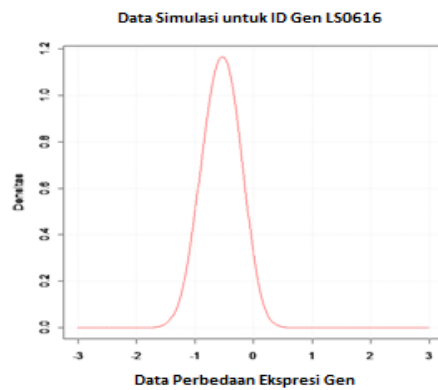


(b)

**Gambar 6.12.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0024

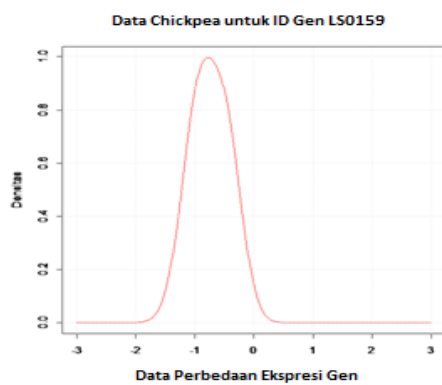


(a)

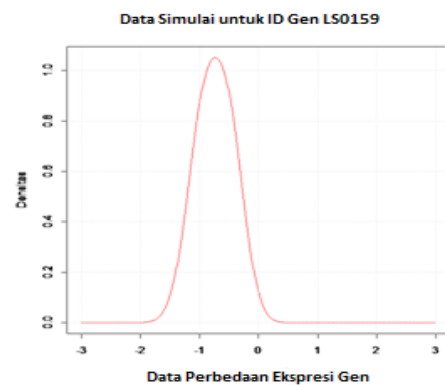


(b)

**Gambar 6.13.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0616

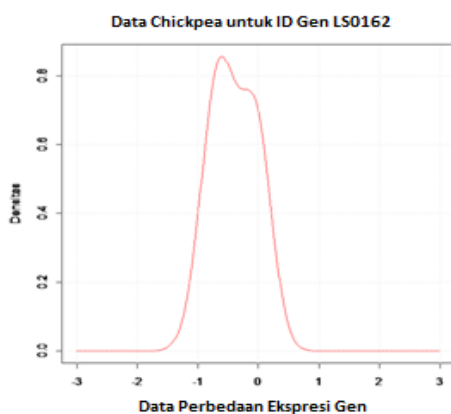


(a)

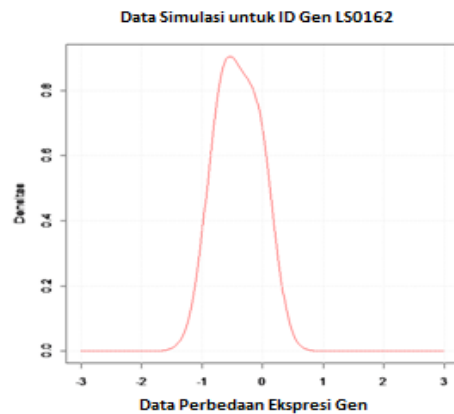


(b)

**Gambar 6.14.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0159

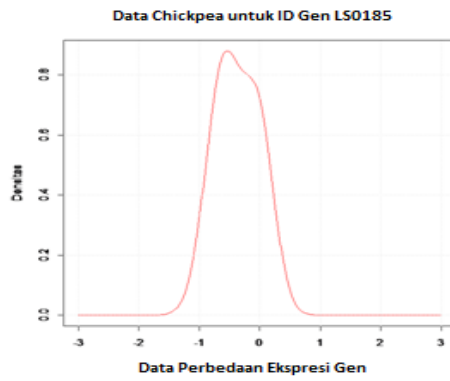


(a)

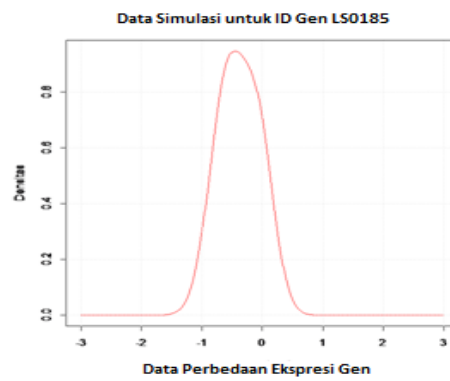


(b)

**Gambar 6.15.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0162

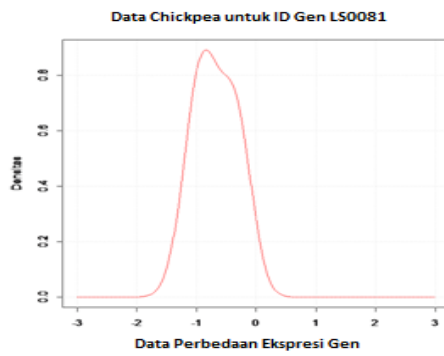


(a)

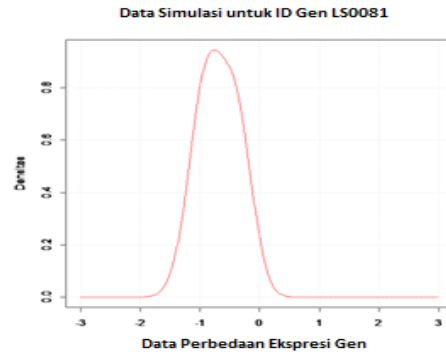


(b)

**Gambar 6.16.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0185

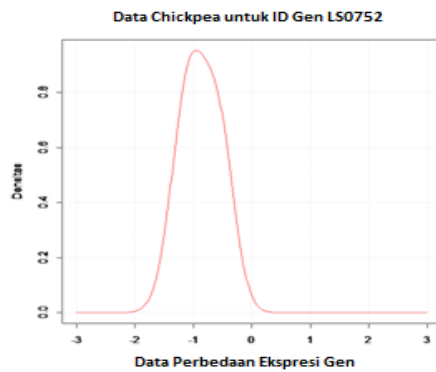


(a)

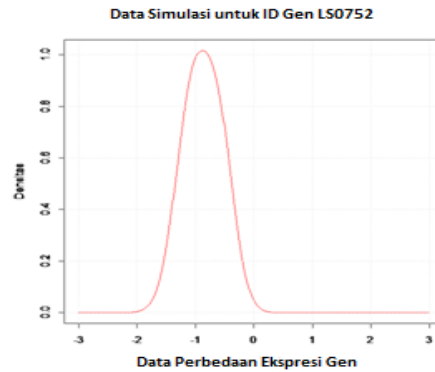


(b)

**Gambar 6.17.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0081

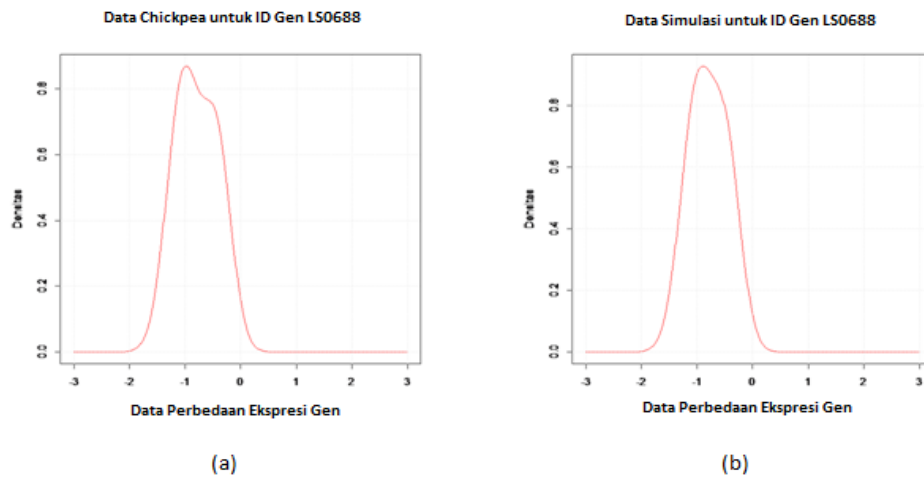


(a)

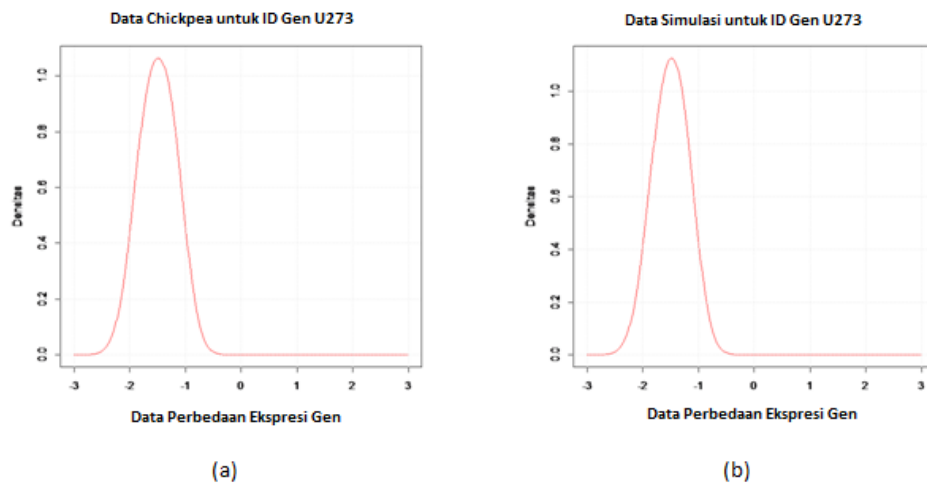


(b)

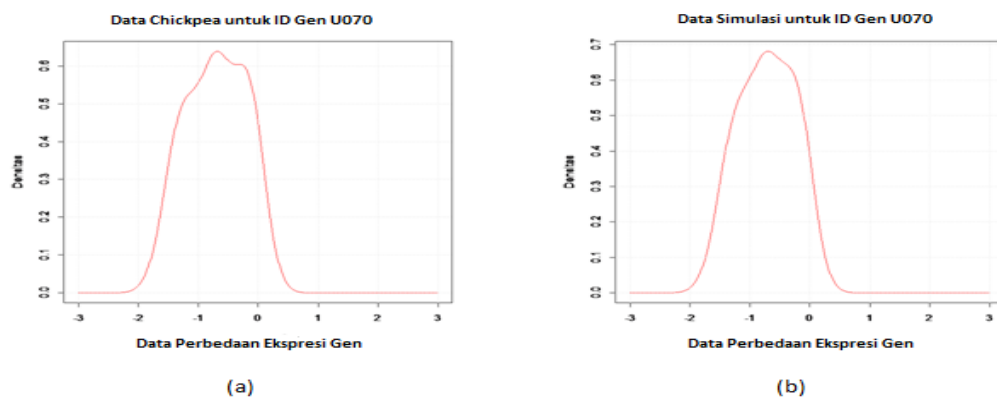
**Gambar 6.18.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0752



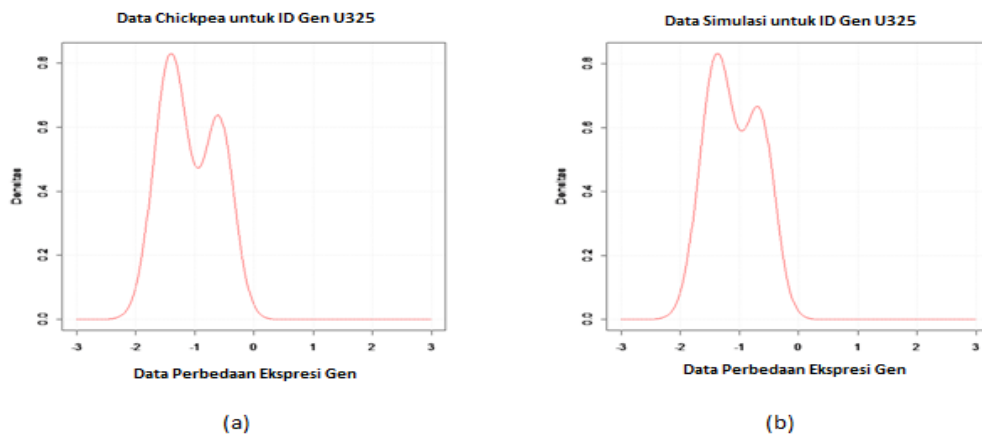
**Gambar 6.19.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0688



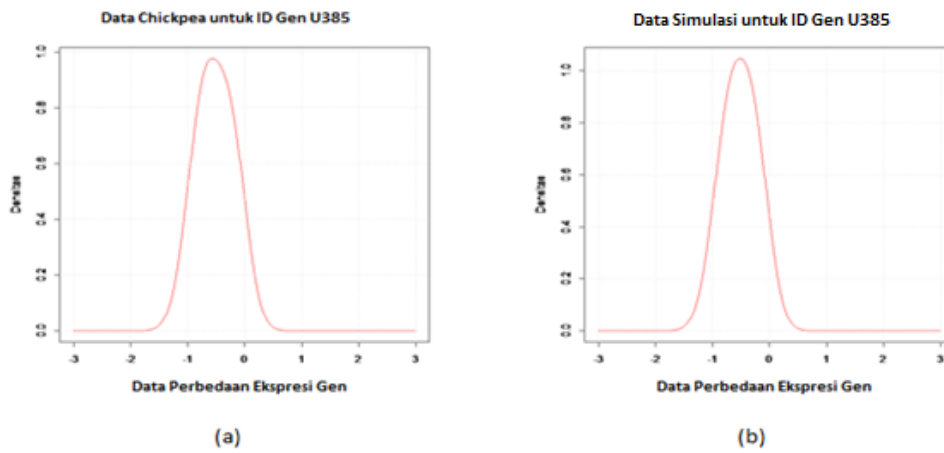
**Gambar 6.20.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U273



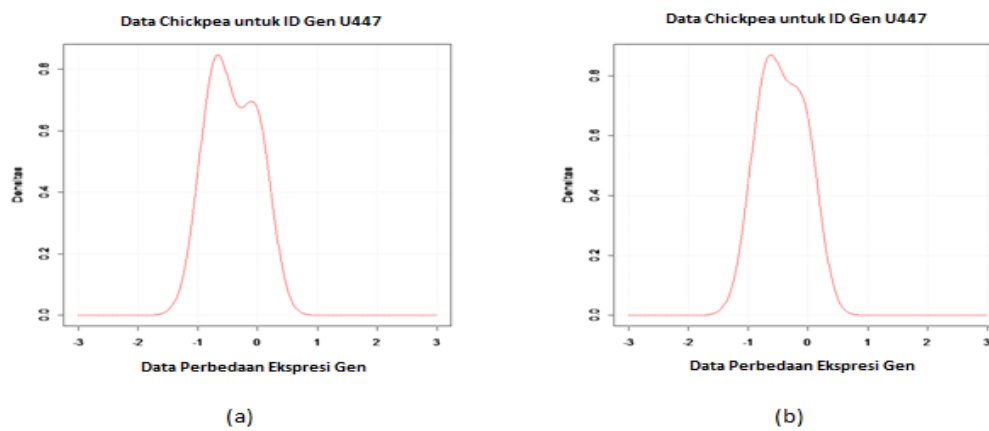
**Gambar 6.21.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U070



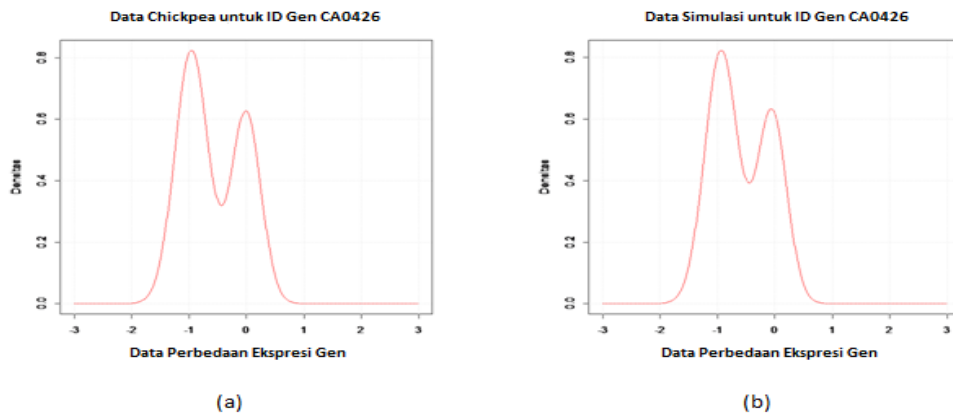
**Gambar 6.22.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U325



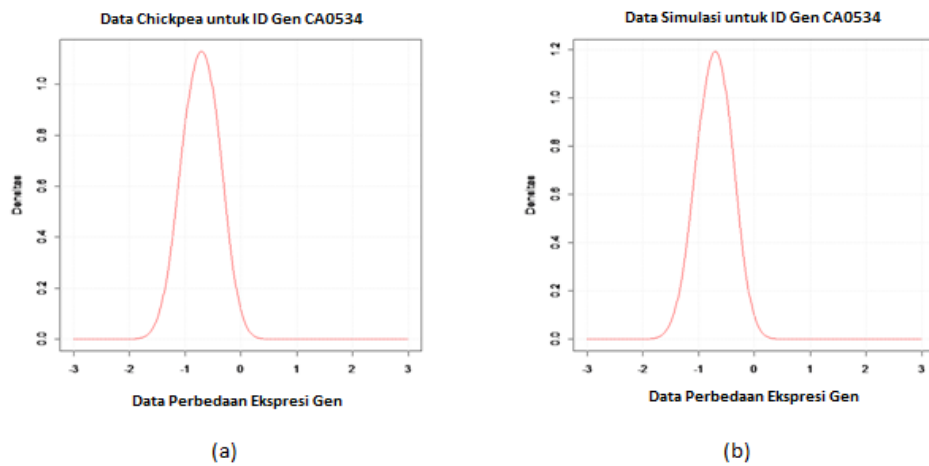
**Gambar 6.23.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U385



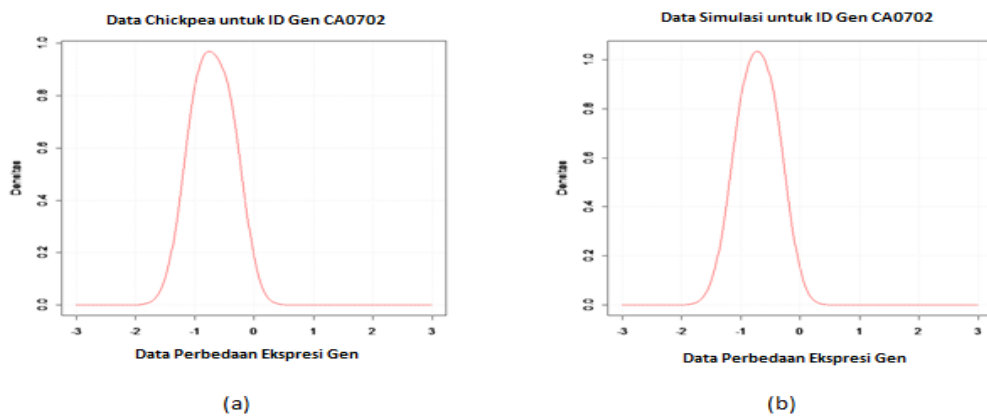
**Gambar 6.24.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U447



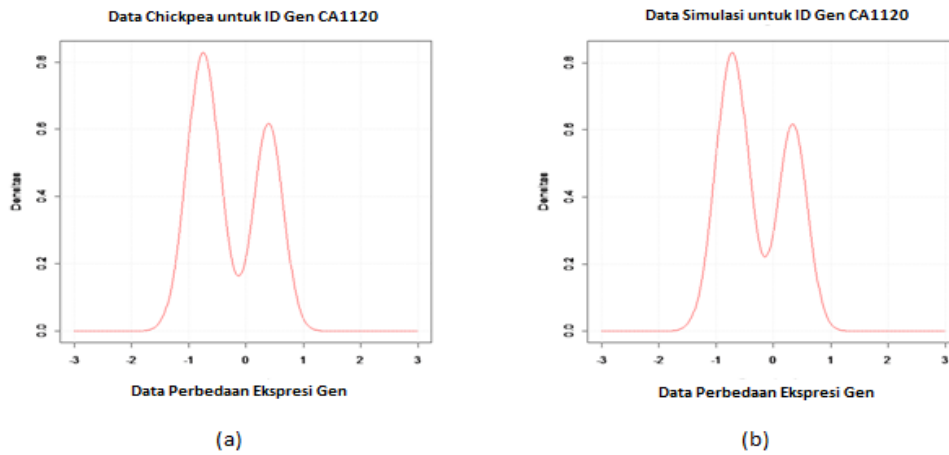
**Gambar 6.25.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0426



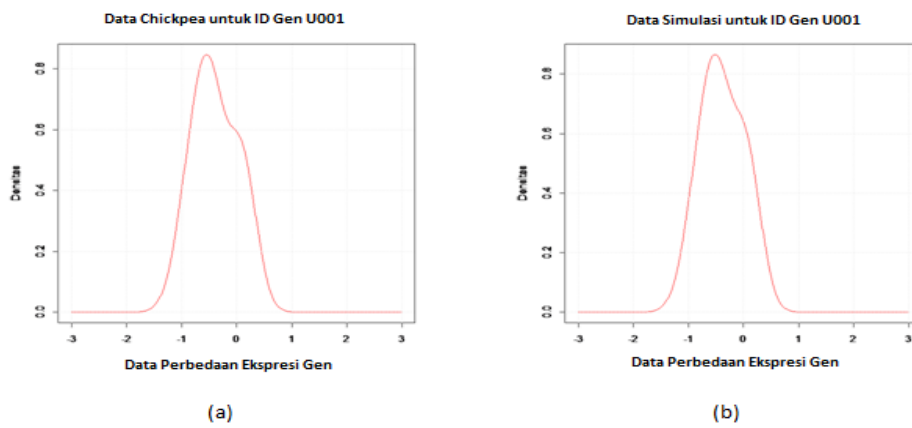
**Gambar 6.26.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0534



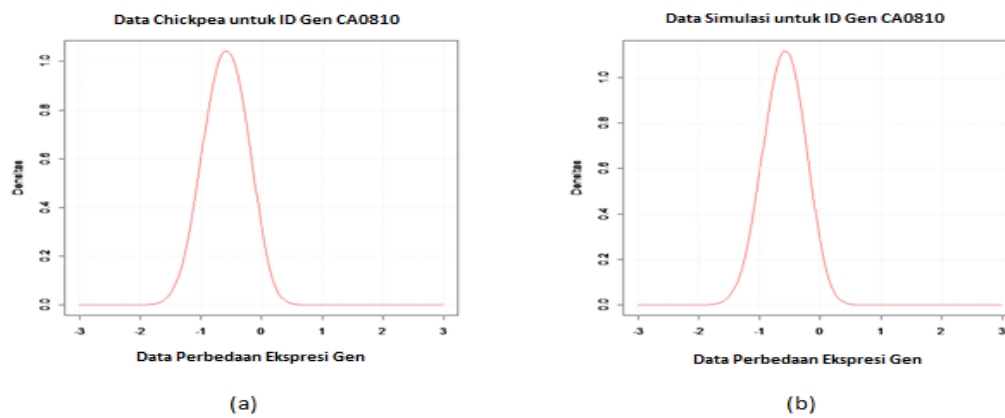
**Gambar 6.27.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0702



**Gambar 6.28.** Bentuk Distribusi BMM-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA1120



**Gambar 6.29.** Bentuk Distribusi BMM-Normal Tiga Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U001



**Gambar 6.30.** Bentuk Distribusi BMM-Normal Tiga Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0810

### 6.3. Implementasi pada Data *Chickpea*

Implementasi Algoritma 6.1 dilakukan pada data *Chickpea* untuk membentuk model BMM-normal, sehingga dapat diketahui sejauhmana kebaikan model yang diterapkan dalam data *Chickpea*. Tahapan-tahapan pemodelan diuraikan sebagai berikut:

#### a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian

Berdasarkan hasil uji pengembangan Algoritma RJMCMC dalam Algoritma 4.3 untuk data *Chickpea*, teridentifikasi satu ID gen dari fungsi ketahanan memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen U070 dan sembilan ID gen memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, dan U273. Sedangkan ID gen dari fungsi energi, ada dua ID gen yang memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen U001 dan CA0810, dan tujuh ID gen memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: U325, U385, U447, CA0426, CA0534, CA0702, dan CA1120. Dalam hal ini berarti bahwa dari 26 ID gen data *Chickpea*, ada tiga ID gen yang memiliki distribusi *mixture* normal tiga komponen dan 16 ID gen yang memiliki distribusi *mixture* normal dua komponen. Hal ini berarti total ada 19 ID gen yang akan dimodelkan *mixture* normal dengan pendekatan Bayesian (BMM-normal) menggunakan implementasi Algoritma 6.1.

Estimasi parameter pada model BMM-normal dilakukan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) dengan distribusi *mixture* normal dua dan tiga komponen. Implementasi Algoritma 2.1 menggunakan program WinBUGS dengan membuat *Syntax BUGS* atau *Graphical Model* (*Doodle BUGS*). *Graphical Model* dan *Syntax BUGS* model BMM-normal untuk distribusi *mixture* normal dua dan tiga komponen sebagaimana yang telah disusun dan digunakan dalam Gambar 6.1 dan Gambar 6.2 serta *Syntax* 6.1 dan *Syntax* 6.2 bagi 19 ID gen benar telah teridentifikasi memiliki distribusi *mixture* normal.

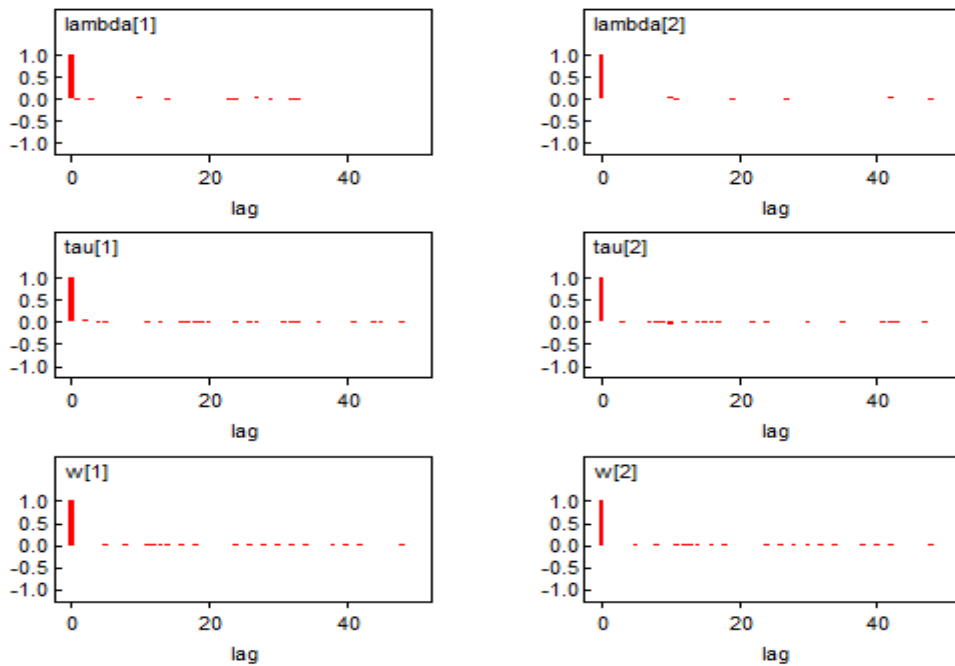
Hasil *running* program BUGS, diberikan contoh ID gen LS0024 untuk *mixture* normal dua komponen (disajikan dalam Tabel 6.5 dan Gambar 6.31



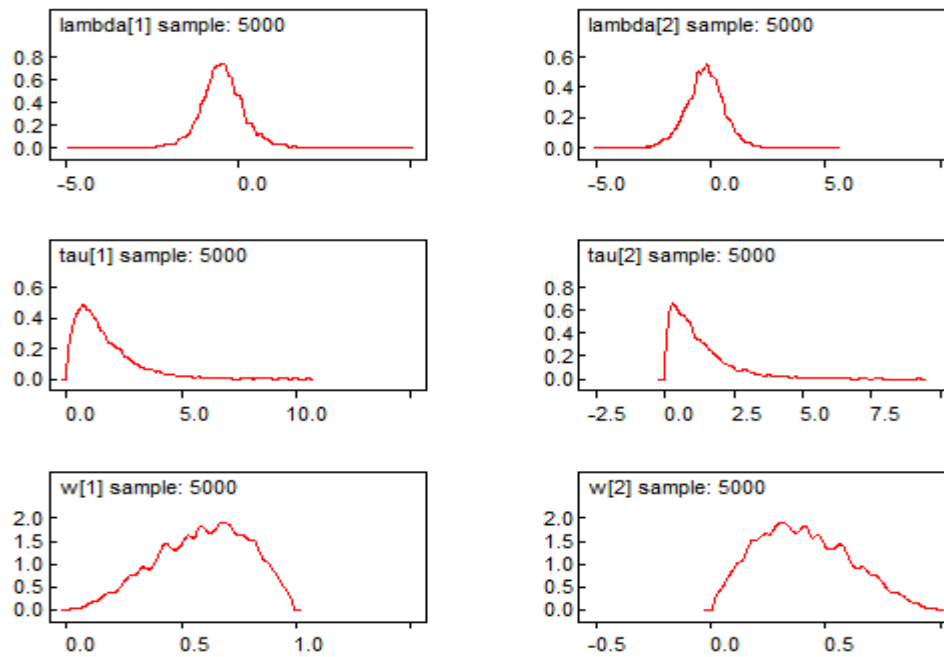
sampai dengan Gambar 6.34) dan ID gen U070 untuk *mixture* normal tiga komponen (disajikan dalam Tabel 6.6 dan Gambar 6.35 sampai dengan Gambar 6.38). Hasil yang disajikan merupakan rata-rata data *Chickpea* untuk pembangkitan data sebanyak 1.000 kali dimana iterasi setiap parameter dilakukan sebanyak 5.000 kali. Hasil secara lengkap untuk 19 ID Gen yang lain, yaitu: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 sebagai pembawa fungsi ketahanan, serta ID gen U325, U385, U447, CA0426, CA0534, CA0702, CA1120, U001, dan CA0810 sebagai pembawa fungsi energi dapat dilihat pada Lampiran 6.

**Tabel 6.5.** Estimasi Parameter Rata-Rata Data *Chickpea* untuk ID Gen LS0024 pada Model BMM-Normal Dua Komponen

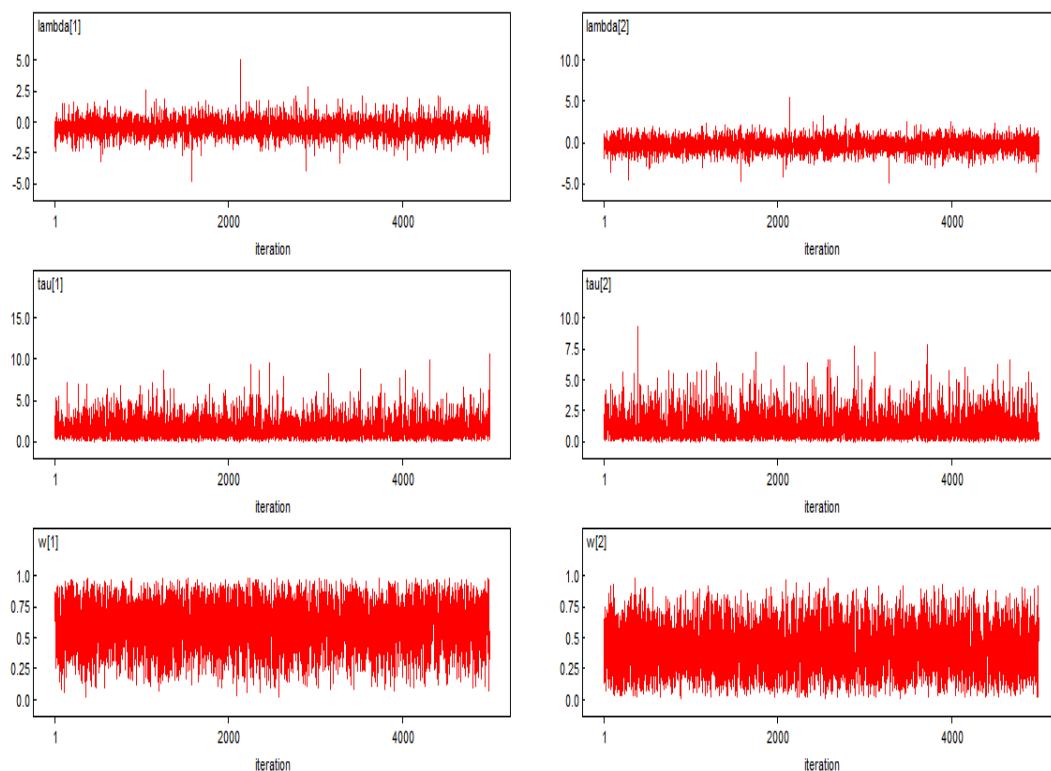
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,4599	0,6467	0,0079	-1,725	-0,4804	0,9101	1	5.000
lambda[2]	-0,2676	0,8347	0,0111	-2,015	-0,2446	1,344	1	5.000
tau[1]	1,574	1,223	0,0177	0,1564	1,259	4,724	1	5.000
tau[2]	1,194	1,038	0,0134	0,0825	0,9014	4,016	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000



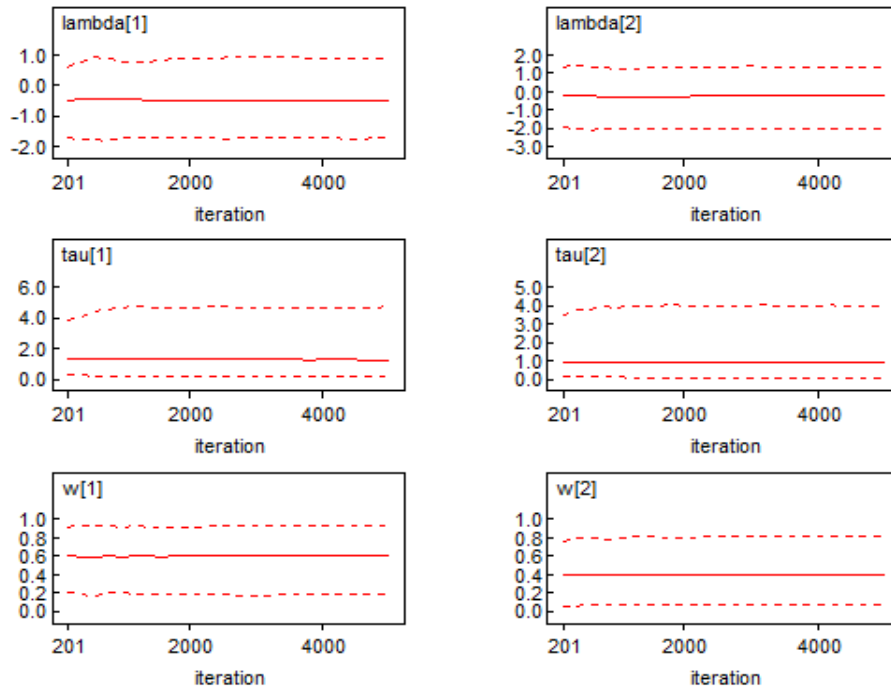
**Gambar 6.31.** Fungsi Autokorelasi setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data *Chickpea*



**Gambar 6.32.** Densitas Kernel setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data *Chickpea*



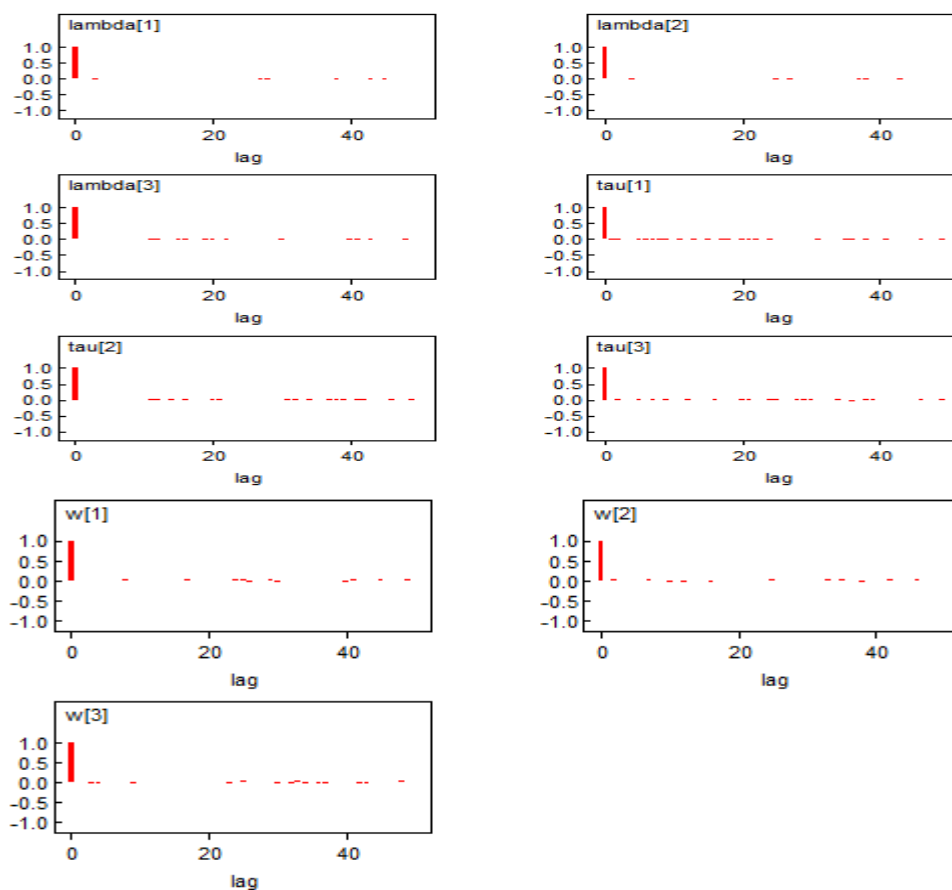
**Gambar 6.33.** History Bangkitan Data Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data *Chickpea*



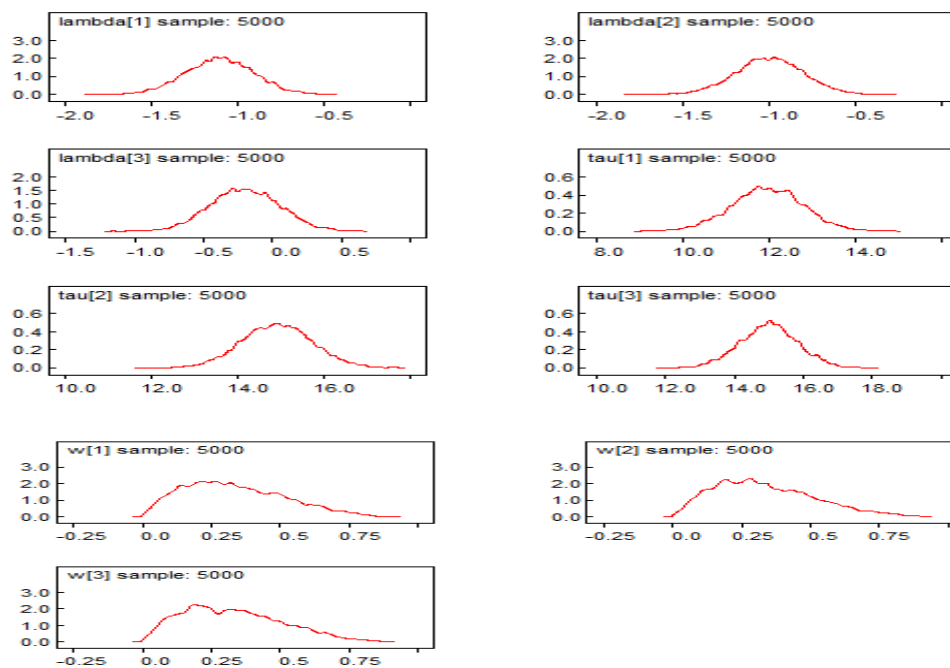
**Gambar 6.34.** *Quantiles* setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0024 Data *Chickpea*

**Tabel 6.6.** Estimasi Parameter Rata-Rata Data *Chickpea* untuk ID Gen U070 pada Model BMM-Normal Tiga Komponen

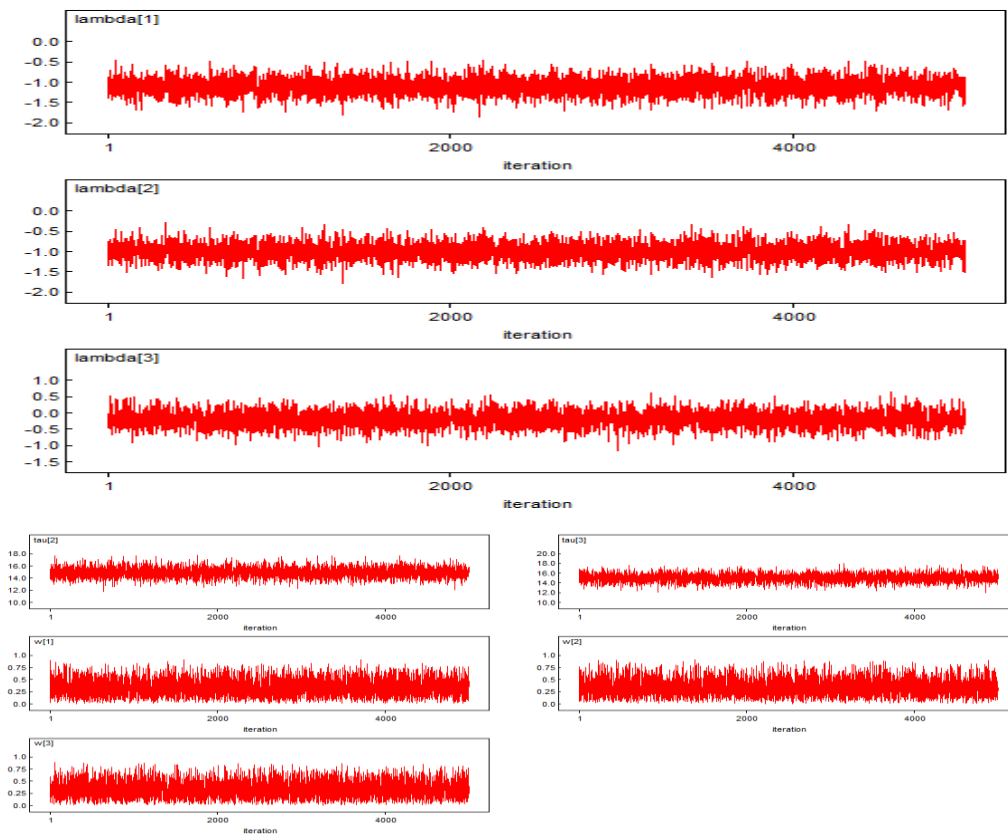
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,075	0,2262	0,0034	-1,526	-1,074	-0,6381	1	5.000
lambda[2]	-0,9481	0,2287	0,0031	-1,381	-0,9517	-0,4758	1	5.000
lambda[3]	-0,1865	0,2566	0,0044	-0,6748	-0,189	0,3329	1	5.000
tau[1]	11,48	3,365	0,0491	6,024	11,05	18,9	1	5.000
tau[2]	13,15	3,73	0,0452	6,934	12,72	2,41	1	5.000
tau[3]	15,12	3,895	0,0573	8,546	14,72	23,72	1	5.000
w[1]	0,3307	0,1782	0,0023	0,0530	0,311	0,7081	1	5.000
w[2]	0,3346	0,1788	0,0025	0,0546	0,3156	0,7162	1	5.000
w[3]	0,3347	0,1757	0,0025	0,0578	0,3173	0,7155	1	5.000



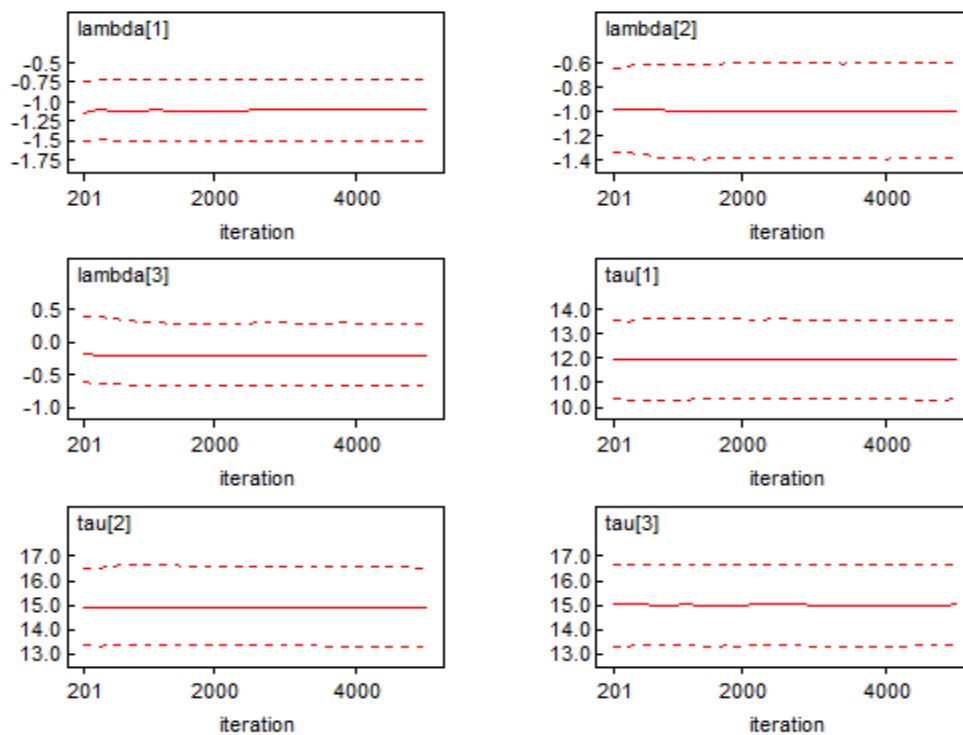
**Gambar 6.35.** Fungsi Autokorelasi setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070 Data *Chickpea*

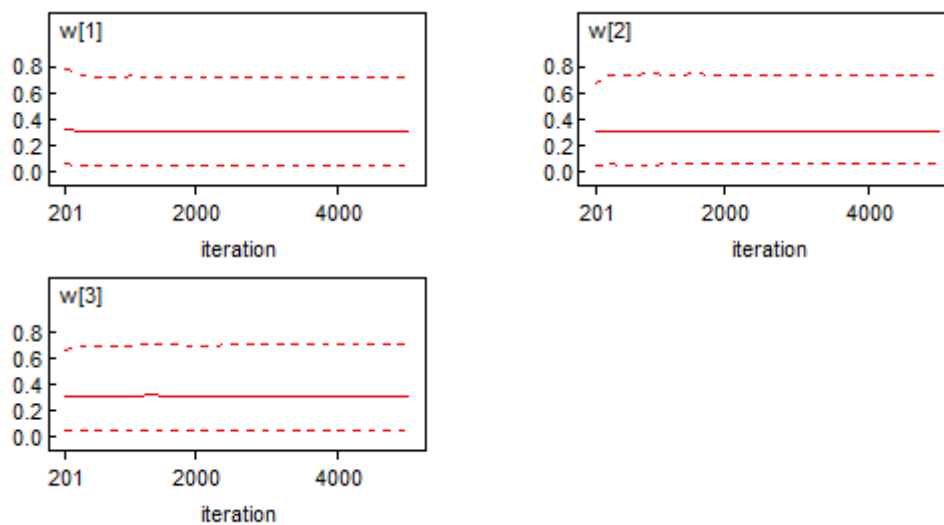


**Gambar 6.36.** Densitas Kernel setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070 Data *Chickpea*



**Gambar 6.37.** History Data Bangkitan Model BMM-Normal Tiga Komponen untuk ID Gen U070 Data *Chickpea*





**Gambar 6.38.** *Quantiles* setiap Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070 Data *Chickpea*

Berdasarkan hasil evaluasi estimasi parameter untuk model BMM-normal dua dan tiga komponen dapat diketahui bahwa untuk 19 ID gen data *Chickpea* yang diamatisudah sangat baik dan sudah memenuhi asumsi pemodelan pendekatan Bayesian, yaitu *independence*, *irreducible*, *aperiodic* dan *recurrent* dalam sampel MCMC untuk mencapai konvergensi estimasinya. Hal ini dapat diketahui dari plot fungsi autokorelasi semua parameter yang diestimasi mempunyai nilai nyata berbeda dengan nol untuk lag-0 dan lag-lag selain lag-0 tidak nyata berbeda dengan nol (Gambar 6.31 dan Gambar 6.35). Demikian juga dengan plot distribusi Kernel, *history*, dan *quantiles* dari parameter yang diestimasi memiliki ciri-ciri yang sesuai dengan model yang diharapkan yaitu estimasi parameter telah mencapai konvergen dengan memenuhi sifat *irreducible*, *aperiodic* dan *recurrent* (Gambar 6.32 sampai dengan Gambar 6.34 dan Gambar 6.36 sampai dengan Gambar 6.38). Oleh karena itu, model BMM-normal dua dan tiga komponen terbaik untuk 19 ID gen data *Chickpea* telah diketahui.

#### **b. Validasi Model BMM-Normal**

Untuk memvalidasi model BMM-Normal yang telah terbentuk digunakan *goodness of fit* KS. Hasil selengkapnya untuk 19 ID gen, yaitu: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 sebagai pembawa fungsi ketahanan serta ID gen: U325, U385, U447, CA0426, CA0534,

CA0702, CA1120, U001, dan CA0810 sebagai pembawa fungsi energi disajikan dalam Tabel 6.7.

Berdasarkan hasil validasi model data simulasi dengan *goodness of fit* KS dapat diketahui bahwa model BMM-normal dua dan tiga komponen mampu memodelkan 19 ID gen yang diamati dengan baik. Hal ini dapat diketahui dari nilai KS yang secara umum memiliki nilai cukup kecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel} = 0,708$ . Model OLS sebagaimana yang telah dijabarkan pada sub bab 2.2, memiliki nilai KS yang secara umum lebih kecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel} = 0,708$  dibandingkan dengan nilai KS model BMM-normal.

### c. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen

HPD dengan *credible interval* 95% dibentuk untuk menentukan nilai-nilai batas *credible interval* dari model yang dibangun, sehingga dapat diketahui pengelompokan dari 19 ID gen yang diamati dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% diperoleh dengan mengimplementasikan Algoritma 5.1. Hasil pembentukan HPD selengkapnya disajikan dalam Tabel 6.8.

Berdasarkan Tabel 6.8 dapat diketahui bahwa ada 10 ID gen yang masuk group *Upregulated*, yaitu enam ID gen dari fungsi ketahanan, ID gen: LS0159, LS0081, LS0752, LS0688, U273, dan U070 serta empat ID gen dari fungsi energi, ID gen: U325, U385, CA0534, dan CA0702. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai negatif (-). Sedangkan sembilan ID gen yang lain masuk group *Regulated*, yaitu empat ID gen dengan fungsi ketahanan: LS0024, LS0616, LS0162, dan LS0185 serta lima ID gen dengan fungsi energi, ID gen: U447, CA0426, CA1120, U001, dan CA0810. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Sementara tidak satupun ditemukan ID gen yang masuk ke dalam kategori *Down-regulated*. Hasil yang diperoleh dengan pemodelan BMM-normal dengan dua dan tiga komponen teridentifikasi 10 ID Gen: LS0159,

LS0081, LS0752, LS0688, U273, U070, U325, U385, CA0534, dan CA0702 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia.

#### **d. Bentuk Distribusi Model BMM-Normal Dua dan Tiga Komponen**

Bentuk distribusi model BMM-normal dua dan tiga komponen data *Chickpea* untuk 19 ID gen yang telah dibangun disajikan secara lengkap dalam Gambar 6.7 sampai dengan Gambar 6.25 sebagaimana yang telah disajikan sebelumnya dimana bentuk distribusi data simulasi mengikuti distribusi data *Chickpea* dan bentuk distribusi BMM-normal setiap ID gen data *Chickpea* adalah sama yaitu berdistribusi *mixture* normal tetapi memiliki perbedaan nilai estimasi parameternya. Bentuk distribusi model BMM-normal dua dan tiga komponen dari data *Chickpea* disajikan secara bersamaan dengan data simulasi karena data simulasi dibangun berdasarkan pola dari data *Chickpea*.



**Tabel 6.7.** *Goodness of Fit* KS Model BMM-Normal Dua dan Tiga Komponen dan Model OLS untuk Data *Chickpea*

No	ID Gen	Nilai Kolmogorov-Smirnov (KS)	
		Model OLS	BMM-Normal
Fungsi Ketahanan:			
1	LS0024	0,2534	0,2590
2	LS0616	0,3800	0,3325
3	LS0159	0,2929	0,2546
4	LS0162	0,2692	0,2707
5	LS0185	0,2752	0,2750
6	LS0081	0,3427	0,3195
7	LS0752	0,3584	0,3226
8	LS0688	0,2554	0,2525
9	U273	0,3415	0,2985
10	U070	0,2291	0,2742
Fungsi Energi:			
11	U325	0,3147	0,2982
12	U385	0,2662	0,2550
13	U447	0,3178	0,3114
14	CA0426	0,2890	0,2599
15	CA0534	0,3318	0,3527
16	CA0702	0,2743	0,2746
17	CA1120	0,3656	0,3381
18	U001	0,2263	0,2679
19	CA0810	0,1855	0,3191

$$KS_{tabel} = 0,708$$

**Tabel 6.8.** HPD dengan *Credible Interval* 95% untuk Model BMM-Normal dengan Dua dan Tiga Komponen Data *Chickpea*

No	ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BMM-Normal Data <i>Chickpea</i>		Hasil Pengelompokan
		Batas Bawah	Batas Atas	
Fungsi Ketahanan:				
1	LS0024	-1,2229	0,3566	<i>Regulated</i>
2	LS0616	-1,4163	0,1121	<i>Regulated</i>
3	LS0159	-1,3187	-0,0189	<i>Up-regulated</i>
4	LS0162	-1,2510	0,2183	<i>Regulated</i>
5	LS0185	-1,5069	0,2971	<i>Regulated</i>
6	LS0081	-1,9190	-0,4255	<i>Up-regulated</i>
7	LS0752	-1,4643	-0,0644	<i>Up-regulated</i>
8	LS0688	-1,7314	-0,1113	<i>Up-regulated</i>
9	U273	-1,9716	-0,8877	<i>Up-regulated</i>
10	U070	-1,5533	-0,0833	<i>Up-regulated</i>
Fungsi Energi:				
11	U325	-2,1925	-0,5276	<i>Up-regulated</i>
12	U385	-1,4816	-0,0611	<i>Up-regulated</i>
13	U447	-0,9976	0,3731	<i>Regulated</i>
14	CA0426	-1,3793	0,7864	<i>Regulated</i>
15	CA0534	-1,3711	-0,2240	<i>Up-regulated</i>
16	CA0702	-1,8035	-0,6930	<i>Up-regulated</i>
17	CA1120	-1,0810	1,0074	<i>Regulated</i>
18	U001	-1,2365	0,3634	<i>Regulated</i>
19	CA0810	-1,3607	0,0043	<i>Regulated</i>

*Halaman ini sengaja dikosongkan*

## BAB VII

### IDENTIFIKASI PERBEDAAN EKSPRESI GEN TANAMAN *CHICKPEA* DI INDONESIA DENGAN PEMODELAN BAYESIAN *MIXTURE* MODEL *AVERAGING*

Dalam bab ini akan disajikan beberapa algoritma yang terkait dengan identifikasi perbedaan ekspresi gen tanaman *Chickpea* di Indonesia melalui pemodelan Bayesian *Mixture* Normal *Averaging* (BMMA-normal) dan implementasi beberapa algoritma yang telah dibangun dilakukan pada data simulasi dan data *Chickpea*.

#### 7.1. Pemodelan dan Algoritma untuk Bayesian *Mixture* Model *Averaging* (BMMA) Normal

Pemodelan Bayesian *Mixture Averaging* adalah sebuah pemodelan yang menggunakan pendekatan Bayesian yang mana pemodelan ini merupakan gabungan dari pemodelan Bayesian *Mixture* (BMM) dan Bayesian *Averaging* (BMA) yang telah dibahas pada bab-bab sebelumnya. Pemodelan BMMA dibentuk dengan asumsi bahwa dalam distribusi data memiliki modus lebih dari satu (multimodal), dimana data akan memiliki susunan sub-sub populasi atau group-group populasi dimana setiap sub populasi merupakan komponen penyusun dari model *mixture* yang memiliki proporsi berbeda-beda. Perbedaanannya dengan pemodelan BMM adalah pemodelan BMMA ini membentuk model tunggal terbaik dengan mempertimbangkan semua kemungkinan model yang bisa dibentuk, sehingga ketidakpastian model dapat diselesaikan. Melalui prinsip tersebut maka pemodelan BMMA dilakukan dengan merata-ratakan distribusi posterior dari semua model terbaiknya.

Dalam pemodelan BMMA-normal diawali dengan pembentukan model BMM-normal dengan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) sebagaimana yang telah dijabarkan pada Bab VI dan dilanjutkan dengan pembentukan model dengan prinsip pemodelan BMA sebagaimana yang telah dijabarkan pada Bab V.

Pada Algoritma 2.1 yang diterapkan dalam data berdistribusi *mixture*, parameter  $\theta$  yang dimaksud adalah  $\lambda[i]$ ,  $\tau[i]$ , dan  $w[i]$ , dengan  $i=1,2,\dots,k$  dimana  $k$  merupakan jumlah komponen dalam distribusi *mixture* normal. Penerapan dari Algoritma 2.1 pada pemodelan BMMA-normal akan memperoleh penduga-penduga parameter untuk  $\lambda[i]$ ,  $\tau[i]$ , dan  $w[i]$  untuk  $\forall i$  dari sebanyak 1.000 model terbaik yang dibangkitkan dengan iterasi masing-masing parameter yang diduga dilakukan sebanyak 5.000 kali. Selanjutnya mengevaluasi 1.000 model BMM-normal yang diperoleh untuk mengetahui pemenuhan asumsi model dengan pendekatan Bayesian. Apabila telah diperoleh 1.000 model terbaik berdasarkan program WinBUGS maka langkah selanjutnya menyeleksi 1.000 model terbaik tersebut dengan metode seleksi model Occam's Window untuk membentuk model BMMA-normal dengan menggunakan prinsip pemodelan BMA-normal. Selanjutnya memvalidasi model BMMA-normal yang terbentuk dengan kriteria *goodness of fit* KS dan menentukan nilai batas bawah dan batas atas dari model BMMA-normal terbaik melalui HPD dengan *credible interval* 95%. Langkah terakhir, menetapkan kelompok data dalam *Down-regulated*, *Regulated*, dan *Up-regulated*. Proses pemodelan BMMA-normal dapat dilakukan menggunakan Algoritma 7.1.

#### **Algoritma 7.1.** Algoritma BMMA-Normal

Tahapan-tahapan dalam Algoritma BMMA-normal melalui langkah-langkah berikut:

- Langkah 1. Bentuk model BMM-normal sebanyak 1.000 model yang dibangkitkan dengan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 5.1) dengan jumlah iterasi setiap parameter yang akan diduga dari model sebanyak 5.000 kali.
- Langkah 2. Evaluasi pemenuhan asumsi pemodelan Bayesian (*independence*, *irreducible*, *aperiodic*, dan *recurrent*) dari 1.000 model BMM-normal yang terbaik melalui kriteria plot fungsi autokorelasi, distribusi Kernel, *history*, dan *quantiles* untuk setiap parameter yang diduga dalam model.

- Langkah 3. Seleksi 1.000 model yang dihasilkan pada Langkah 2 dengan metode seleksi model Occam's Window (Algoritma 5.3) untuk membentuk model BMMA-normal.
- Langkah 4. Menghitung nilai *goodness of fit* KS untuk validasi model BMMA-normal sebagaimana yang telah dibahas dalam sub bab 2.8.
- Langkah 5. Tentukan nilai batas bawah (namakan a) dan batas atas (namakan b) dengan Algoritma 5.1 tentang pembentukan HPD dengan *credible interval* 95%
- Langkah 6. Menetapkan kelompok data ke dalam tiga kelompok, yaitu *Down-regulated*, *Regulated*, dan *Up-regulated*.

Implementasi Algoritma 7.1 (Algoritma BMMA-normal) terhadap data simulasi dan data *Chickpea* telah dilakukan oleh Astuti dkk. (2015c; 2016c). Konsep kerja Algoritma 7.1 juga telah diimplementasi pada data Diabetes Mellitus oleh Astuti dkk. (2016a; 2016b).

Syarat-syarat metode BMMA dapat digunakan jika data memiliki distribusi *mixture* yang akan dimodelkan melalui pendekatan Bayesian dengan metode pendugaan parameter model menggunakan konsep *averaging*. Dalam hal ini, tentu saja persyaratan-persyaratan pemodelan Bayesian *mixture* dan konsep *averaging* sebagaimana yang telah dijabarkan pada sub bab 2.3, sub bab 2.4 dan sub bab 2.7 harus terpenuhi. Sedangkan metode BMMA-normal sebagai kasus khusus dari metode BMMA memiliki persyaratan-persyaratan sebagaimana metode BMMA tetapi untuk data berdistribusi *mixture* normal.

## **7.2. Implementasi pada Data Simulasi**

Implementasi pemodelan BMMA-normal pada data simulasi dilakukan untuk mengetahui sejauh mana kebaikan model ini untuk data simulasi dan sekaligus untuk mengidentifikasi pengelompokan ID gen ke dalam kelompok *Up-regulated*, *Regulated*, dan *Down-regulated*. Tahapan-tahapan pemodelan diuraikan sebagai berikut:

#### a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian

Berdasarkan data simulasi yang telah dibangun dengan ciri data *Chickpea*, teridentifikasi satu ID gen dari fungsi ketahanan yang memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen U070 dan sembilan ID gen memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, dan U273. Sedangkan ID gen dari fungsi energi, ada dua ID gen yang memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen: U001 dan CA0810 dan tujuh ID gen memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: U325, U385, U447, CA0426, CA0534, CA0702, dan CA1120. Artinya, dari 26 ID gen yang diamati pada data simulasi, ada tiga ID gen yang memiliki distribusi *mixture* normal tiga komponen dan 16 ID gen yang memiliki distribusi *mixture* normal dua komponen, sehingga total ada 19 ID gen yang akan dimodelkan *mixture* normal *averaging* dengan pendekatan Bayesian (BMMA-normal) dengan cara mengimplementasikan Algoritma 7.1.

Estimasi parameter untuk membentuk 1.000 model BMM-normal dengan jumlah iterasi setiap parameter yang diduga dalam model sebanyak 5.000 kali, dilakukan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1). Distribusi datanya adalah *mixture* normal dua dan tiga komponen. Implementasi Algoritma 2.1 ini menggunakan program WinBUGS dengan membuat *Syntax BUGS* atau *Graphical Model (Doodle BUGS)* sebagaimana yang telah ditunjukkan pada Gambar 6.2 dan Gambar 6.3, serta *Syntax* 6.1 dan *Syntax* 6.2, sebagaimana yang telah digunakan sebelumnya pada saat pembentukan model BMM-normal untuk 19 ID gen yang telah teridentifikasi memiliki distribusi *mixture* normal.

Sesuai dengan hasil yang telah diperoleh sebelumnya pada pemodelan BMM-normal, sebagai contoh *running* program BUGS diterapkan pada ID gen LS0024 yang berdistribusi *mixture* normal dengan dua komponen. Hasilnya disajikan dalam Tabel 6.1 dan Gambar 6.4 sampai dengan Gambar 6.7. Program ini juga diterapkan pada ID gen U070 yang berdistribusi *mixture* normal tiga. Hasilnya disajikan dalam Tabel 6.2 dan Gambar 6.8 sampai dengan Gambar 6.11. Hasil-hasil tersebut merupakan rata-rata data simulasi untuk pembangkitan

data 1.000 kali dimana iterasi setiap parameter yang diduga dalam model dilakukan sebanyak 5.000 kali. Hasil selengkapnya untuk 19 ID gen yang lain, yaitu LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 sebagai pembawa fungsi ketahanan, serta ID gen: U325, U385, U447, CA0426, CA0534, CA0702, CA1120, U001, dan CA0810 sebagai pembawa fungsi energi dapat dilihat pada Lampiran 5.

Hasil evaluasi estimasi parameter untuk model BMM-normal dua dan tiga komponen sesuai pembahasan sebelumnya telah didapatkan bahwa untuk 19 ID gen dari data simulasi yang diamatisudah sangat baik dan sudah memenuhi asumsi pemodelan pendekatan Bayesian, yaitu *independence*, *irreducible*, *aperiodic* dan *recurrent* dalam sampel MCMC untuk mencapai konvergensi estimasinya. Hal ini dapat diketahui dari plot fungsi autokorelasi semua parameter yang diestimasi mempunyai nilai nyata berbeda dengan nol untuk lag-0 dan lag-lag selain lag-0 tidak nyata berbeda dengan nol (Gambar 6.4 dan Gambar 6.8). Demikian juga dengan plot distribusi Kernel, *history*, dan *quantiles* dari parameter yang diestimasi memiliki ciri-ciri yang sesuai dengan model yang diharapkan, yaitu estimasi parameter telah mencapai konvergen dengan memenuhi sifat *irreducible*, *aperiodic* dan *recurrent* (Gambar 6.5 sampai dengan Gambar 6.7 dan Gambar 6.9 sampai dengan Gambar 6.11). Oleh karena itu, model BMM-normal dua dan tiga komponen terbaik untuk 19 ID gen data simulasi telah diketahui.

#### **b. Seleksi Model dengan Metode Occam's Window**

Langkah 3 dari Algoritma 7.1 melakukan seleksi model untuk 1.000 model BMM-normal yang telah dibangun untuk membentuk model BMMA-normal. Seleksi untuk 1.000 model BMM-normal terbaik dilakukan dengan metode seleksi Occam's Window dengan mengimplementasikan Algoritma 5.2, sehingga dapat diperoleh model BMMA-normal terbaik. Hasil yang diperoleh disajikan secara lengkap pada Tabel 7.1.

Berdasarkan hasil seleksi Occam's Window pada Tabel 7.1 dapat diketahui bahwa pada pemodelan BMMA-normal dua dan tiga komponen harus dibangun dengan melibatkan 916 model yang terseleksi untuk setiap ID gen yang diamati. Secara khusus rata-rata model yang terlibat dalam pemodelan BMMA-



normal dua komponen sebanyak 930 model dan untuk rata-rata model yang terlibat dalam pemodelan BMMA-normal tiga komponen sebanyak 855 model. Model BMMA-normal dua dan tiga komponen yang terbangun cukup kompleks, tetapi dengan perkembangan kemajuan teknologi komputasi saat ini, maka kompleksitas model bukan menjadi masalah dalam estimasi modelnya.

**Tabel 7.1.** Hasil Seleksi Model dengan Metode Occam's Window untuk Data Simulasi

No	ID Gen	Rata-Rata Hasil Seleksi 1.000 Model BMM-Normal dengan Metode Occam's Window untuk Data Simulasi dalam Pemodelan BMMA-Normal dari Pengulangan 1.000 kali
Fungsi Ketahanan:		
1	LS0024	933,64
2	LS0616	934,52
3	LS0159	936,57
4	LS0162	933,02
5	LS0185	933,07
6	LS0081	932,62
7	LS0752	932,28
8	LS0688	933,65
9	U273	937,17
10	U070	800,33
Fungsi Energi:		
11	U325	922,89
12	U385	936,10
13	U447	930,35
14	CA0426	915,45
15	CA0534	929,12
16	CA0702	932,23
17	CA1120	918,73
18	U001	875,35
19	CA0810	888,87
Rata-Rata		<b>915,74</b>

### c. Validasi Model BMMA-Normal

Untuk memvalidasi model BMMA-normal yang telah terbentuk digunakan *goodness of fit* KS. Hasil selengkapnya untuk 19 ID gen, yaitu: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 sebagai pembawa fungsi ketahanan, serta ID gen: U325, U385, U447, CA0426, CA0534,

CA0702, CA1120, U001, dan CA0810 sebagai pembawa fungsi energi, disajikan dalam Tabel 7.2.

Berdasarkan hasil validasi model data simulasi dengan *goodness of fit* KS dalam Tabel 7.2 dapat diketahui bahwa model BMMA-normal dua dan tiga komponen mampu memodelkan 19 ID gen yang diamati dengan baik. Hal ini dapat diketahui dari nilai rata-rata KS yang secara umum memiliki nilai cukup kecil dan masih dalam jaminan lebih kecil dari  $KS_{tabel} = 0,708$ . Nilai rata-rata KS model BMMA-normal ini lebih kecil dan masih dalam jaminan lebih kecil dari  $KS_{tabel} = 0,708$  dibandingkan dengan model OLS sebagaimana yang telah dijabarkan dalam sub bab 2.2.

**Tabel 7.2.** *Goodness of Fit* KS Model BMMA-Normal dengan Dua dan Tiga Komponen dan Model OLS Data Simulasi

No	ID Gen	Nilai Rata-Rata Kolmogorov-Smirnov (KS) Dari 1.000 Kali Pembangkitan Data	
		Model OLS	BMMA-Normal
Fungsi Ketahanan:			
1	LS0024	0,2670	0,2745
2	LS0616	0,3479	0,3159
3	LS0159	0,3058	0,2883
4	LS0162	0,2793	0,2673
5	LS0185	0,2877	0,2824
6	LS0081	0,3465	0,3167
7	LS0752	0,3417	0,3299
8	LS0688	0,2637	0,2631
9	U273	0,3477	0,3176
10	U070	0,2344	0,2179
Fungsi Energi:			
11	U325	0,3275	0,2997
12	U385	0,2774	0,2774
13	U447	0,3318	0,3007
14	CA0426	0,2986	0,6433
15	CA0534	0,2896	0,3917
16	CA0702	0,2892	0,3108
17	CA1120	0,3684	0,3319
18	U001	0,2304	0,2466
19	CA0810	0,2328	0,3233

$KS_{tabel} = 0,708$

#### **d. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen**

HPD dengan *credible interval* 95% dibentuk untuk menentukan nilai-nilai batas *credible interval*, sehingga dapat diketahui pengelompokan dari 19 ID gen yang diamati ke dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% diperoleh dengan mengimplementasikan Algoritma 5.1. Hasil pembuatan HPD selengkapnya disajikan dalam Tabel 7.3.

Berdasarkan Tabel 7.3 dapat diketahui bahwa ada tujuh ID gen yang masuk group *Up-regulated*, empat ID gen dari fungsi ketahanan, yaitu ID gen: LS0159, LS0752, LS0688, dan U273, serta tiga ID gen dari fungsi energi, yaitu ID gen: U325, CA0534, dan CA0702. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Sedangkan 12 ID gen yang lain masuk group *Regulated*, yaitu enam ID gen dengan fungsi ketahanan: LS0024, LS0616, LS0162, LS0185, LS0081 dan U070, serta enam ID gen dengan fungsi energi, yaitu ID gen: U385, U447, CA0426, CA1120, U001, dan CA0810. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Sementara itu tidak ada satupun ID gen yang masuk dalam kategori *Down-regulated*. Hasil yang diperoleh dengan pemodelan BMMA-normal dua dan tiga komponen teridentifikasi tujuh ID Gen: LS0159, LS0752, LS0688, U273, U325, CA0534, dan CA0702 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia.

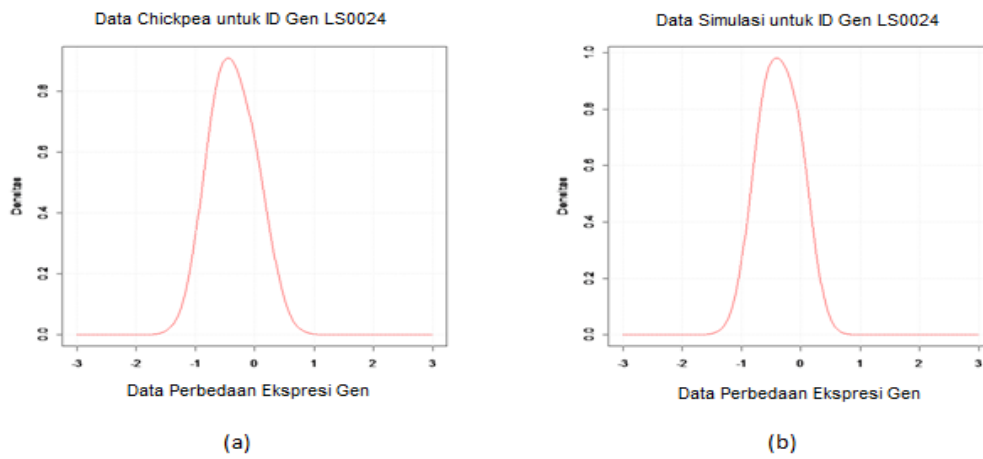
#### **e. Bentuk Distribusi Model BMMA-Normal Dua dan Tiga Komponen**

Bentuk distribusi BMMA-normal data simulasi disajikan bersamaan dengan data *Chickpea*. Hal ini dilakukan karena data simulasi dibangun berdasarkan pola data *Chickpea*. Bentuk distribusi model BMMA-normal dengan dua dan tiga komponen dari data *Chickpea* dan data simulasi untuk 19 ID gen yang telah dibangun disajikan secara lengkap dalam Gambar 7.1 sampai dengan Gambar 7.19. Sedangkan beberapa contoh model persamaan BMMA dan model persamaan pendekatan OLS untuk berbagai ID gen ditunjukkan di Lampiran 9.

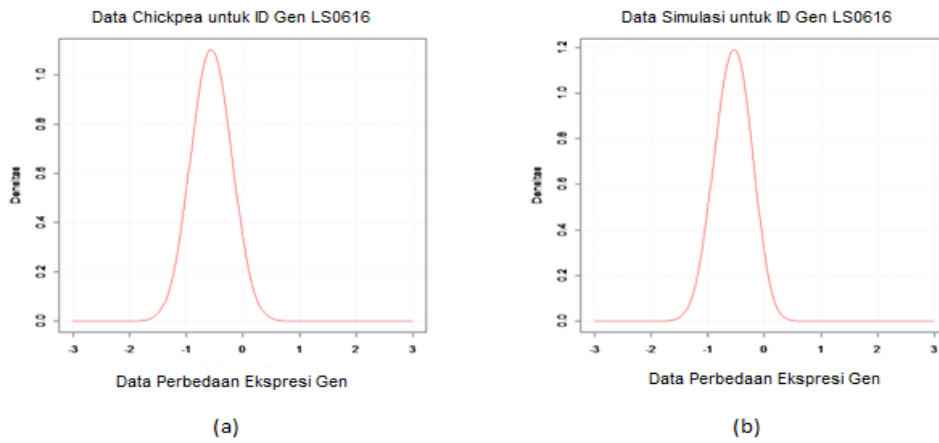
Hasil yang diperoleh menunjukkan bahwa distribusi data simulasi mengikuti distribusi data *Chickpea* dan bentuk distribusi setiap ID gen adalah sama yaitu berdistribusi *mixture* normal tetapi berbeda-beda nilai estimasi parameternya.

**Tabel 7.3.** HPD dengan *Credible Interval* 95% untuk Model BMMA-Normal dengan Dua dan Tiga Komponen Data Simulasi

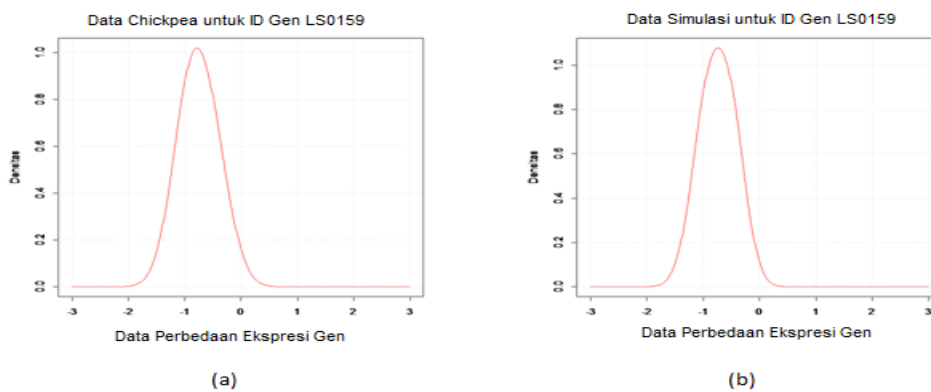
No	ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BMMA-Normal Data Simulasi		Hasil Pengelompokan
		Batas Bawah	Batas Atas	
Fungsi Ketahanan:				
1	LS0024	-1,1186	0,3922	<i>Regulated</i>
2	LS0616	-1,2325	0,1088	<i>Regulated</i>
3	LS0159	-1,4649	-0,0314	<i>Up-regulated</i>
4	LS0162	-1,1899	0,3798	<i>Regulated</i>
5	LS0185	-1,1420	0,3782	<i>Regulated</i>
6	LS0081	-1,4490	0,0741	<i>Regulated</i>
7	LS0752	-1,6065	-0,1480	<i>Up-regulated</i>
8	LS0688	-1,5644	-0,0222	<i>Up-regulated</i>
9	U273	-2,1967	-0,8222	<i>Up-regulated</i>
10	U070	-1,8287	0,2263	<i>Regulated</i>
Fungsi Energi:				
11	U325	-1,9574	-0,1932	<i>Up-regulated</i>
12	U385	-1,2450	0,1919	<i>Regulated</i>
13	U447	-1,2392	0,3850	<i>Regulated</i>
14	CA0426	-1,5096	0,4160	<i>Regulated</i>
15	CA0534	-1,3888	-0,0707	<i>Up-regulated</i>
16	CA0702	-1,4505	-0,0051	<i>Up-regulated</i>
17	CA1120	-1,2955	0,7995	<i>Regulated</i>
18	U001	-1,2812	0,4803	<i>Regulated</i>
19	CA0810	-1,3465	0,0734	<i>Regulated</i>



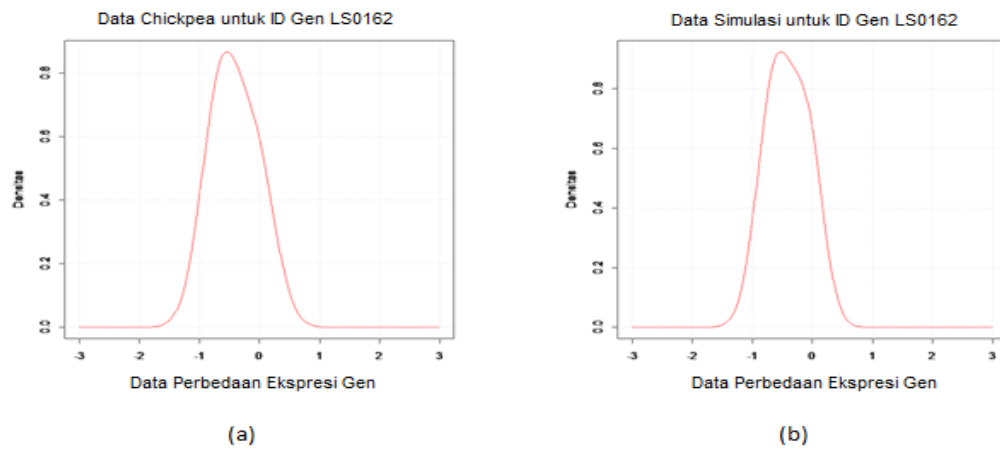
**Gambar 7.1.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0024



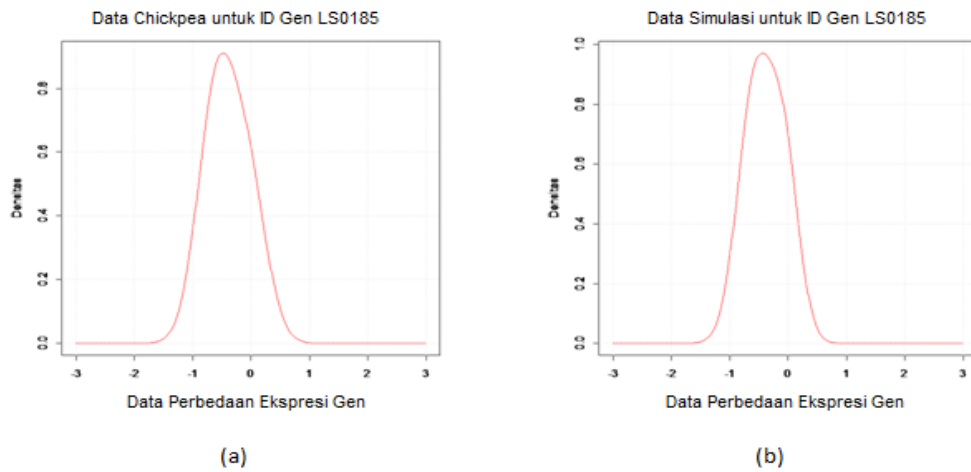
**Gambar 7.2.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0616



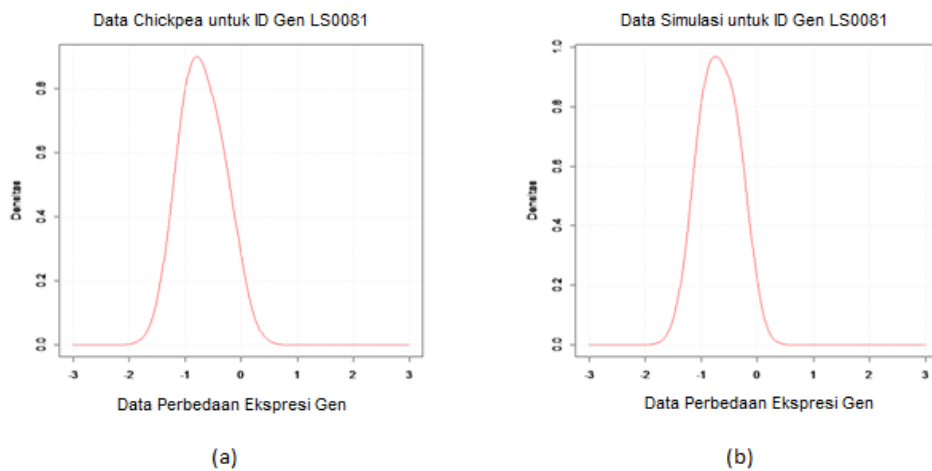
**Gambar 7.3.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0159



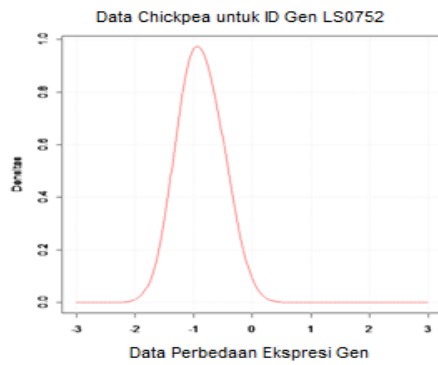
**Gambar 7.4.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0162



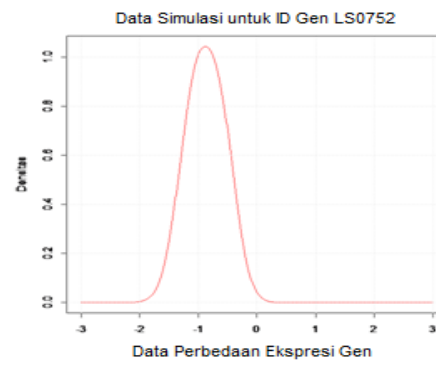
**Gambar 7.5.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0185



**Gambar 7.6.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0081

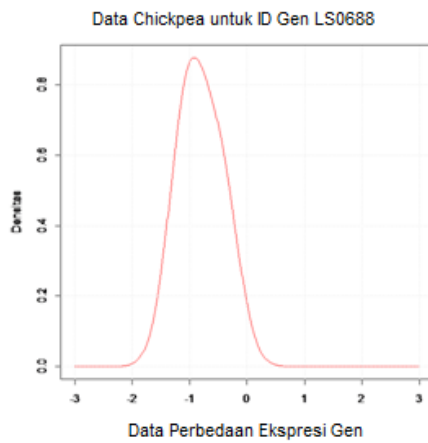


(a)

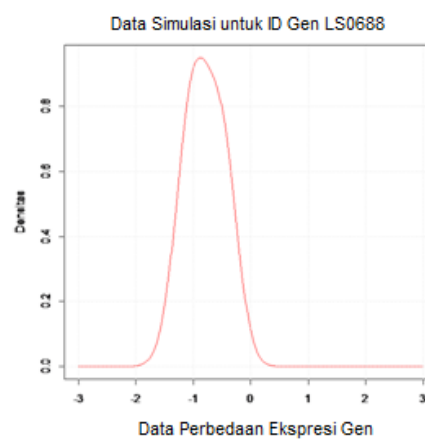


(b)

**Gambar 7.7.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0752

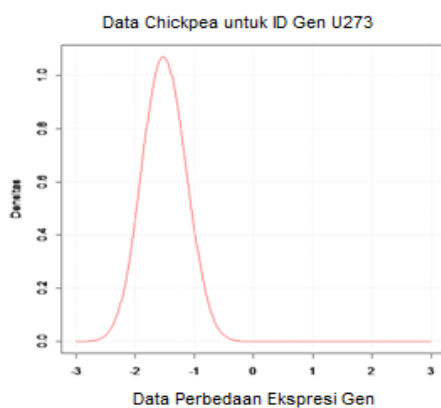


(a)

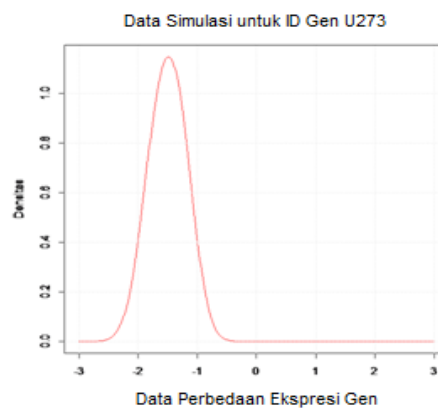


(b)

**Gambar 7.8.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen LS0688

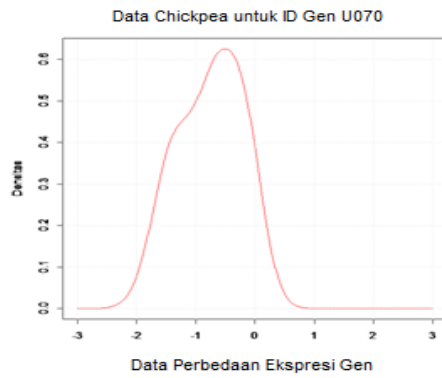


(a)

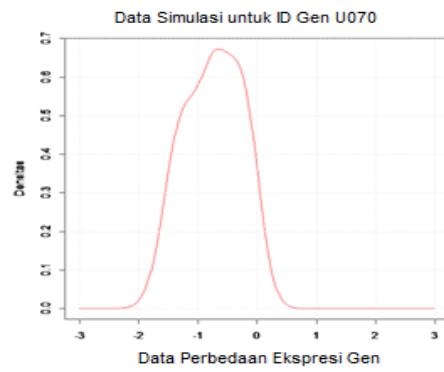


(b)

**Gambar 7.9.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U273

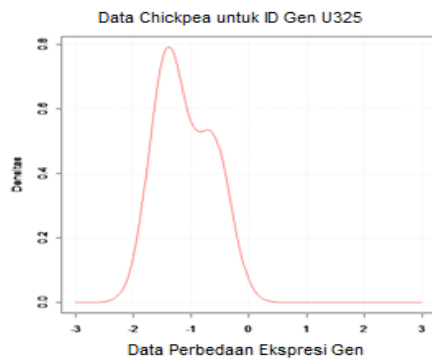


(a)

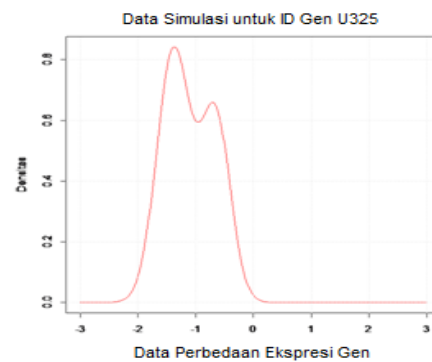


(b)

**Gambar 7.10.** Bentuk Distribusi BMMA-Normal Tiga Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U070

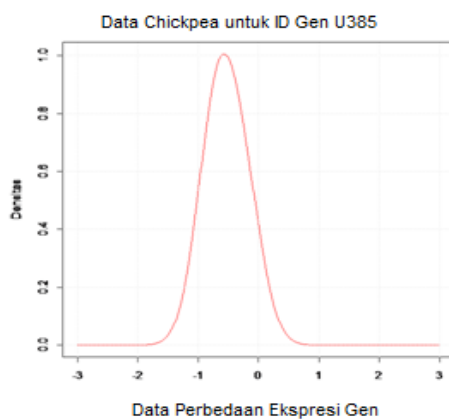


(a)

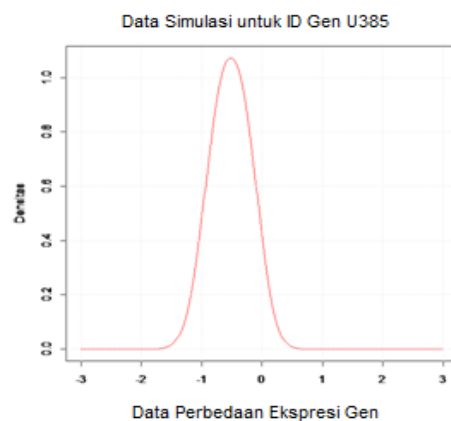


(b)

**Gambar 7.11.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U325



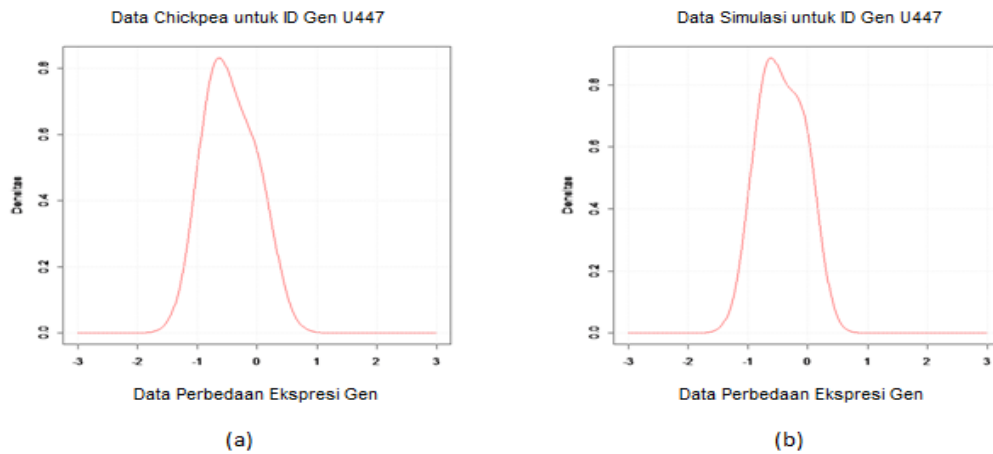
(a)



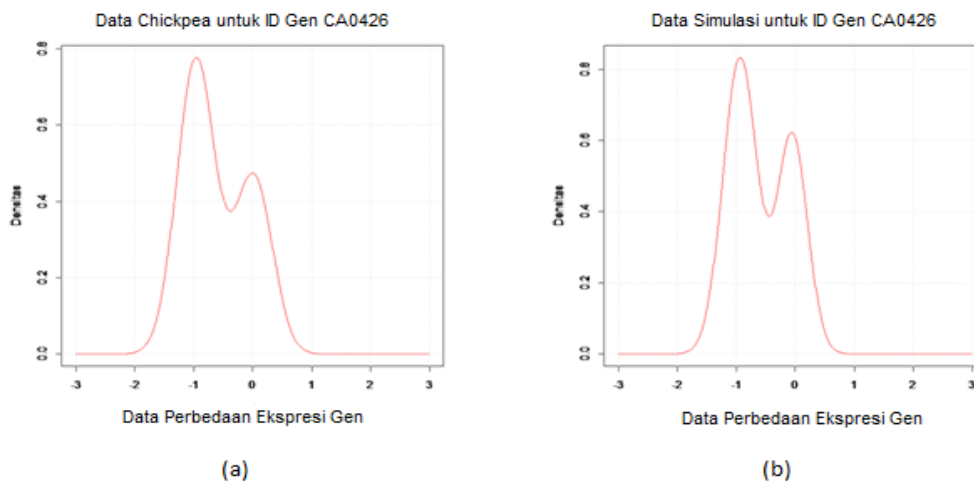
(b)

**Gambar 7.12.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U385

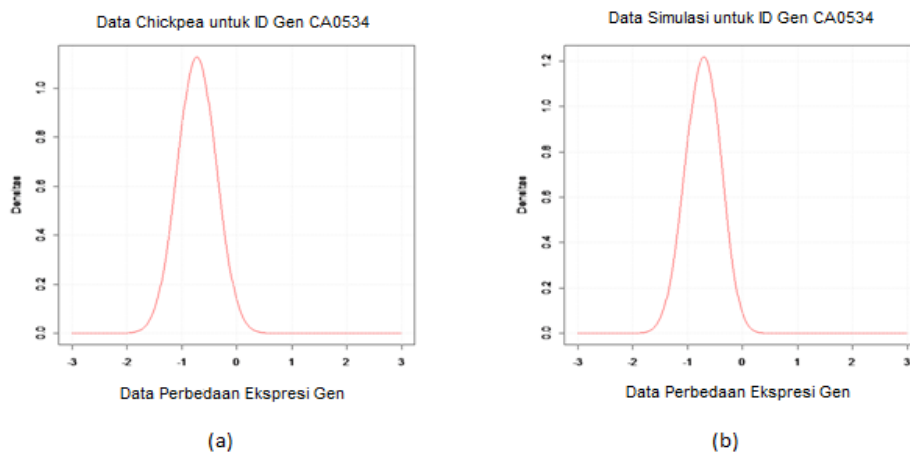




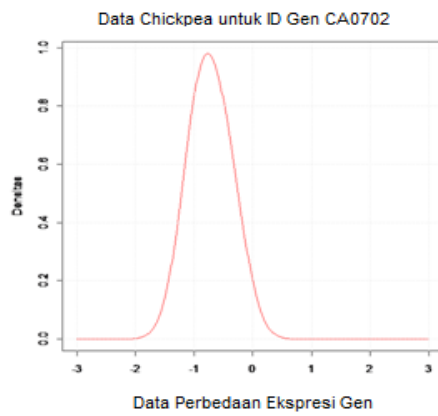
**Gambar 7.13.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U447



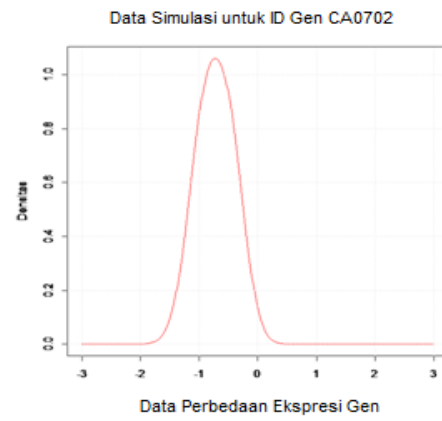
**Gambar 7.14.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0426



**Gambar 7.15.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0534

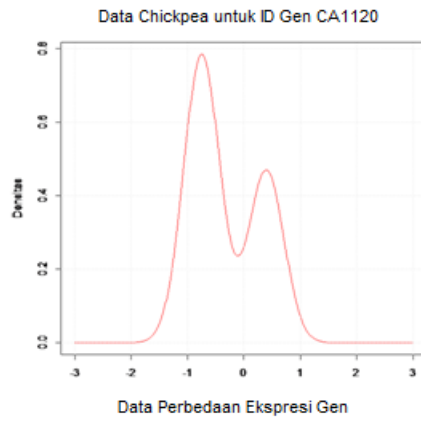


(a)

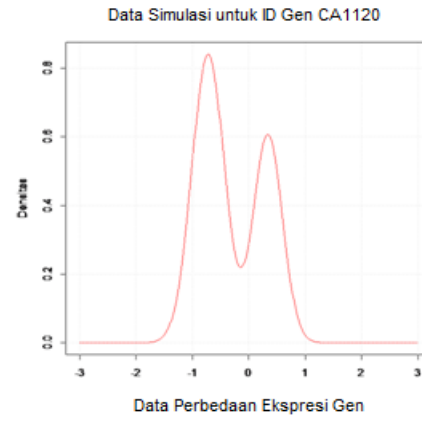


(b)

**Gambar 7.16.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0702

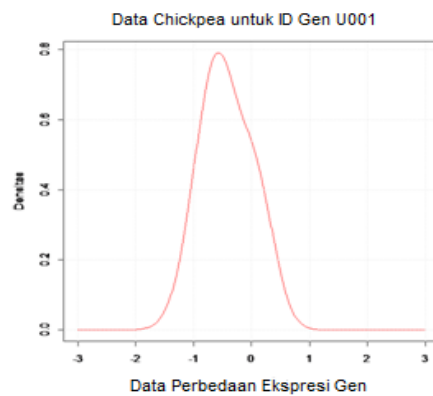


(a)

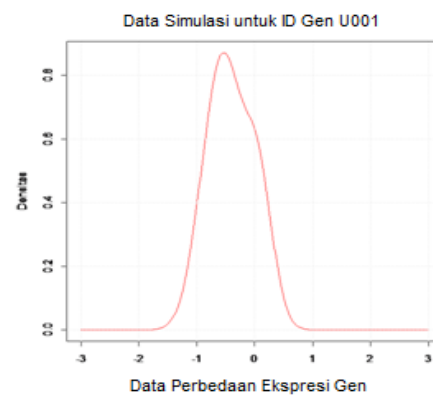


(b)

**Gambar 7.17.** Bentuk Distribusi BMMA-Normal Dua Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA1120

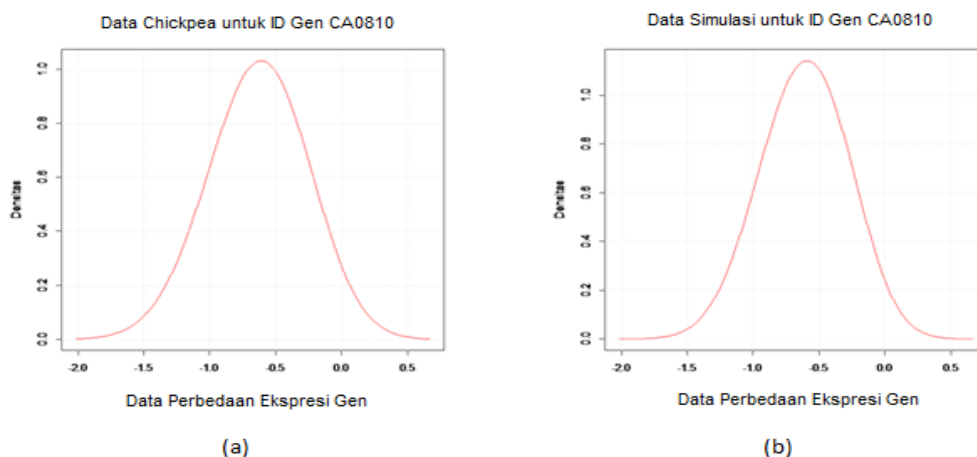


(a)



(b)

**Gambar 7.18.** Bentuk Distribusi BMMA-Normal Tiga Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen U001



**Gambar 7.19.** Bentuk Distribusi BMMA-Normal Tiga Komponen: (a) Data *Chickpea* dan (b) Data Simulasi untuk ID Gen CA0810

### 7.3. Implementasi pada Data *Chickpea*

Implementasi pemodelan pada data *Chickpea* dilakukan untuk mengetahui sejauhmana kebaikan model BMMA-normal yang diperoleh, serta mengidentifikasi pengelompokan data *Chickpea* ke dalam kelompok *Up-regulated*, *Regulated*, dan *Down-regulated*. Tahapan-tahapan pemodelan diuraikan sebagai berikut:

#### a. Estimasi Parameter dan Pemenuhan Asumsi Model Bayesian

Berdasarkan hasil uji pengembangan Algoritma RJMCMC dalam Algoritma 4.3 untuk data *Chickpea*, teridentifikasi satu ID gen dari fungsi ketahanan yang memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen: U070 dan sembilan ID gen yang memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, dan U273. Sedangkan ID gen dari fungsi energi, ada dua ID gen yang memiliki distribusi *mixture* normal tiga komponen, yaitu ID gen: U001 dan CA0810, dan tujuh ID gen memiliki distribusi *mixture* normal dua komponen, yaitu ID gen: U325, U385, U447, CA0426, CA0534, CA0702, dan CA1120. Artinya, dari 26 ID gen data *Chickpea*, ada tiga ID gen yang memiliki distribusi *mixture* normal tiga komponen dan 16 ID gen yang memiliki distribusi *mixture* normal dua komponen. Hal ini berarti total ada 19 ID gen yang akan dimodelkan

*mixture* normal dengan pendekatan Bayesian (BMM-normal) menggunakan implementasi Algoritma 6.1.

Estimasi parameter pada model BMM-normal dilakukan menggunakan Algoritma MCMC pendekatan *Gibbs Sampler* (Algoritma 2.1) dengan distribusi *mixture* normal dua dan tiga komponen. Implementasi Algoritma 2.1 menggunakan program WinBUGS dengan membuat *Syntax BUGS* atau *Graphical Model* (*Doodle BUGS*). *Graphical Model* dan *Syntax BUGS* model BMM-normal untuk distribusi *mixture* normal dua dan tiga komponen sebagaimana yang telah disusun dan digunakan dalam Gambar 6.1 dan Gambar 6.2 serta *Syntax* 6.1 dan *Syntax* 6.2 bagi 19 ID gen benar telah teridentifikasi memiliki distribusi *mixture* normal.

Hasil *running* program BUGS, diberikan contoh ID gen LS0024 untuk *mixture* normal dua komponen (disajikan dalam Tabel 6.5 dan Gambar 6.31 sampai dengan Gambar 6.34) dan ID gen U070 untuk *mixture* normal tiga komponen (disajikan dalam Tabel 6.6 dan Gambar 6.35 sampai dengan Gambar 6.38). Hasil yang disajikan merupakan rata-rata data *Chickpea* untuk pembangkitan data sebanyak 1.000 kali dimana iterasi setiap parameter dilakukan sebanyak 5.000 kali. Hasil secara lengkap untuk 19 ID Gen yang lain, yaitu: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 dengan fungsi ketahanan, serta ID gen U325, U385, U447, CA0426, CA0534, CA0702, CA1120, U001, dan CA0810 dengan fungsi energi dapat dilihat pada Lampiran 6.

Berdasarkan hasil evaluasi estimasi parameter untuk model BMM-normal dua dan tiga komponen dapat diketahui bahwa untuk 19 ID gen dari data *Chickpea* yang diamati sudah sangat baik dan sudah memenuhi asumsi pemodelan pendekatan Bayesian, yaitu *independence*, *irreducible*, *aperiodic* dan *recurrent* dalam sampel MCMC untuk mencapai konvergensi estimasinya. Hal ini dapat diketahui dari plot fungsi autokorelasi semua parameter yang diestimasi mempunyai nilai nyata berbeda dengan nol untuk lag-0 dan lag-lag selain lag-0 tidak nyata berbeda dengan nol (Gambar 6.31 dan Gambar 6.35). Demikian juga dengan plot distribusi Kernel, *history*, dan *quantiles* dari parameter yang diestimasi memiliki ciri-ciri yang sesuai dengan model yang diharapkan, yaitu

estimasi parameter telah mencapai konvergen dengan memenuhi sifat *irreducible*, *aperiodic* dan *recurrent* (Gambar 6.32 sampai dengan Gambar 6.34 dan Gambar 6.36 sampai dengan Gambar 6.38). Oleh karena itu, model BMM-normal dua dan tiga komponen terbaik untuk 19 ID gen data *Chickpea* telah diketahui.

#### **b. Seleksi Model dengan Metode Occam's Window**

Langkah 3 dari Algoritma 7.1 melakukan seleksi model untuk 1.000 model BMM-normal yang telah dibangun untuk membentuk model BMMA-normal. Seleksi untuk 1.000 model BMM-normal terbaik dengan metode seleksi model Occam's Window dilakukan dengan mengimplementasikan Algoritma 5.4 untuk data *Chickpea*, sehingga dapat diperoleh model BMMA-normal terbaik. Hasil yang diperoleh disajikan secara lengkap pada Tabel 7.4.

Berdasarkan hasil seleksi Occam's Window pada Tabel 7.4 dapat diketahui bahwa pada pemodelan BMMA-normal dua dan tiga komponen harus dibangun dengan melibatkan 918 model yang terseleksi untuk setiap ID gen yang diamati. Secara khusus rata-rata model yang terlibat dalam pemodelan BMMA-normal dua komponen sebanyak 930 model dan untuk rata-rata model yang terlibat dalam pemodelan BMMA-normal tiga komponen sebanyak 817 model. Model BMMA-normal dua dan tiga komponen yang terbangun cukup kompleks, tetapi dengan perkembangan kemajuan teknologi komputasi saat ini, maka kompleksitas model bukan menjadi masalah dalam estimasi modelnya.

#### **c. Validasi Model BMMA-Normal**

Untuk memvalidasi model BMMA-normal yang telah terbentuk digunakan *goodness of fit* KS. Hasil selengkapnya untuk 19 ID gen, yaitu: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 sebagai pembawa fungsi ketahanan, serta ID gen U325, U385, U447, CA0426, CA0534, CA0702, CA1120, U001, dan CA0810 sebagai pembawa fungsi energi disajikan dalam Tabel 7.5.

Berdasarkan hasil validasi model data simulasi dengan *goodness of fit* KS dalam Tabel 7.5 dapat diketahui bahwa model BMMA-normal dua dan tiga komponen mampu memodelkan 19 ID gen yang diamati dengan baik. Hal ini

dapat diketahui dari nilai KS yang cukup kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$ . Nilai KS model BMMA-normal ini lebih kecil dan masih dalam jaminan memiliki nilai yang lebih kecil dari  $KS_{tabel} = 0,708$  dibandingkan dengan model OLS sebagaimana yang telah dijabarkan dalam sub bab 2.2.

#### **d. HPD dengan *Credible Interval* 95% dan Pengelompokan ID Gen**

HPD dengan *credible interval* 95% dibentuk untuk menentukan nilai-nilai batas *credible interval*, sehingga diketahui pengelompokan dari 19 ID gen yang diamati ke dalam tiga kelompok, yaitu *Up-regulated*, *Regulated*, dan *Down-regulated*. HPD dengan *credible interval* 95% diperoleh dengan mengimplementasikan Algoritma 2.3. Hasil pembuatan HPD selengkapnya disajikan dalam Tabel 7.6.

Berdasarkan Tabel 7.6 dapat diketahui bahwa ada lima ID gen yang masuk group *Up-regulated*, tiga ID gen dari fungsi ketahanan, yaitu ID gen: LS0159, LS0752, dan U273, serta dua ID gen dari fungsi energi, yaitu ID gen: U385 dan CA0534. Hal ini dapat dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas yang bernilai positif (+). Sedangkan 14 ID gen yang lain masuk group *Regulated*, yaitu tujuh ID gen dengan fungsi ketahanan: LS0024, LS0616, LS0162, LS0185, LS0081, LS0688, dan U070 serta tujuh ID gen dengan fungsi energi, yaitu ID gen: U385, U447, CA0426, CA1120, CA0702, U001, dan CA0810. Group *Regulated* dilihat dari nilai HPD dengan *credible interval* 95% dengan batas bawah bernilai negatif (-) dan batas atas bernilai positif (+). Sementara itu, tidak ada satupun ID gen yang masuk dalam kategori *Down-regulated*. Hasil yang diperoleh dengan pemodelan BMMA-normal dua dan tiga komponen teridentifikasi lima ID Gen: LS0159, LS0752, U273, U325, dan CA0534 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* dan diharapkan mampu meningkatkan produksi tanaman *Chickpea* di dunia. ID gen-ID gen yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* inilah yang dapat digunakan untuk menemukan

varietas baru bagi tanaman *Chickpea* jenis *Lasseter* muda dengan cara penyisipan gen.

**Tabel 7.4.** Hasil Seleksi Model dengan Metode Occam's Window untuk Data *Chickpea*

No	ID Gen	Hasil Seleksi 1.000 Model BMM-Normal dengan Metode Occam's Window untuk Data <i>Chickpea</i> dalam Pemodelan BMMA-Normal
Fungsi Ketahanan:		
1	LS0024	939
2	LS0616	938
3	LS0159	941
4	LS0162	929
5	LS0185	923
6	LS0081	930
7	LS0752	929
8	LS0688	934
9	U273	936
10	U070	817
Fungsi Energi:		
11	U325	928
12	U385	943
13	U447	922
14	CA0426	906
15	CA0534	946
16	CA0702	931
17	CA1120	907
18	U001	861
19	CA0810	883
Rata-Rata		915,74

**Tabel 7.5.** *Goodness of Fit* KS Model BMMA-Normal Dua dan Tiga Komponen dan Model OLS untuk Data *Chickpea*

No	ID Gen	Nilai Kolmogorov-Smirnov (KS)	
		Model OLS	BMMA-Normal
Fungsi Ketahanan:			
1	LS0024	0,2534	0,2422
2	LS0616	0,3800	0,3524
3	LS0159	0,2929	0,2741
4	LS0162	0,2692	0,2588
5	LS0185	0,2752	0,2573
6	LS0081	0,3427	0,3032
7	LS0752	0,3584	0,3211
8	LS0688	0,2554	0,2394
9	U273	0,3415	0,3156
10	U070	0,2291	0,2242
Fungsi Energi:			
11	U325	0,3147	0,2863
12	U385	0,2662	0,2466
13	U447	0,3178	0,2871
14	CA0426	0,2890	0,2471
15	CA0534	0,3318	0,3640
16	CA0702	0,2743	0,2715
17	CA1120	0,3656	0,3241
18	U001	0,2263	0,2349
19	CA0810	0,1855	0,2862

$KS_{tabel} = 0,708$



**Tabel 7.6.** HPD dengan *Credible Interval* 95% untuk Model BMMA-Normal dengan Dua dan Tiga Komponen Data *Chickpea*

No	ID Gen	HPD dengan <i>Credible Interval</i> 95% Model BMMA-Normal Data <i>Chickpea</i>		Hasil Pengelompokan
		Batas Bawah	Batas Atas	
Fungsi Ketahanan:				
1	LS0024	-1,1429	0,4393	<i>Regulated</i>
2	LS0616	-1,2519	0,1393	<i>Regulated</i>
3	LS0159	-1,4834	-0,0042	<i>Up-regulated</i>
4	LS0162	-1,2077	0,4258	<i>Regulated</i>
5	LS0185	-1,1633	0,4231	<i>Regulated</i>
6	LS0081	-1,4763	0,1250	<i>Regulated</i>
7	LS0752	-1,6328	-0,1071	<i>Up-regulated</i>
8	LS0688	-1,5962	0,0188	<i>Regulated</i>
9	U273	-2,2179	-0,7887	<i>Up-regulated</i>
10	U070	-1,8668	0,2715	<i>Regulated</i>
Fungsi Energi:				
11	U325	-1,9771	-0,1483	<i>Up-regulated</i>
12	U385	-1,2658	0,2232	<i>Regulated</i>
13	U447	-1,2707	0,4354	<i>Regulated</i>
14	CA0426	-1,5363	0,4794	<i>Regulated</i>
15	CA0534	-1,4102	-0,0410	<i>Up-regulated</i>
16	CA0702	-1,4723	0,0404	<i>Regulated</i>
17	CA1120	-1,3197	0,8565	<i>Regulated</i>
18	U001	-1,3120	0,5255	<i>Regulated</i>
19	CA0810	-1,3849	0,1092	<i>Regulated</i>

#### e. Bentuk Distribusi Model BMMA-Normal Dua dan Tiga Komponen

Bentuk distribusi model BMMA-normal data *Chickpea* disajikan bersamaan dengan bentuk distribusi model BMMA-normal data simulasi. Hal ini dilakukan karena data simulasi dibangun berdasarkan data *Chickpea*. Bentuk distribusi model BMMA-normal dua dan tiga komponen dari data *Chickpea* untuk 19 ID gen yang telah dibangun disajikan secara lengkap dalam Gambar 7.1 sampai dengan Gambar 7.19 sebagaimana yang telah disajikan sebelumnya. Hasil yang diperoleh menunjukkan bahwa bentuk distribusi data simulasi tetap mengikuti data *Chickpea* dan bentuk distribusi setiap ID gennya adalah sama,

yaitu berdistribusi *mixture* normal tetapi nilai estimasi parameternya yang berbeda-beda.

#### 7.4. Perbandingan Hasil Pemodelan BMM-Normal dan BMMA-Normal

Hasil pemodelan BMM-normal dan BMMA-normal yang telah diperoleh akan dibandingkan kebaikan modelnya berdasarkan nilai *goodness of fit* KS yang telah diperoleh sehingga diketahui model terbaik di antara model BMM-normal dan BMMA-normal. Selain itu juga akan ditunjukkan hasil pengelompokan ID gen yang telah diperoleh berdasarkan nilai batas HPD dengan *credible interval* 95%. Tabel 7.7 menyajikan nilai *goodness of fit* KS model BMM-normal dan BMMA-normal untuk data simulasi dan data *Chickpea* serta hasil pengelompokan ID gen.

Dalam Tabel 7.7 dapat ditunjukkan bahwa model *Bayesian Mixture Model Averaging* normal (BMMA-normal) adalah model yang terbaik dibandingkan dengan model BMM-normal dan model OLS. Hal ini ditunjukkan baik pada data simulasi maupun data *Chickpea*. Kebaikan model ditunjukkan oleh nilai KS dari model BMMA-normal, sebagian besar dari ID gen memiliki nilai terkecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel} = 0,708$  dibandingkan dengan dua model yang lain, yaitu model OLS dan model BMM-normal. Hasil pengelompokan ID gen berdasarkan nilai batas HPD dengan *credible interval* 95% menunjukkan hasil yang berbeda antara model BMM-normal dan model BMMA-normal untuk beberapa ID gen baik untuk data simulasi maupun data *Chickpea*. Perbedaan dan persamaan hasil pengelompokan ID gen dapat dilihat selengkapnya dalam Tabel 7.7.

Berdasarkan hasil yang diperoleh pada penelitian ini telah ditunjukkan bahwa model BMMA-normal adalah model yang paling cocok untuk data *microarray* di Indonesia kasus data *Chickpea*, yaitu data perbedaan ekspresi gen pada kondisi sakit (terserang penyakit jamur patogen *Ascochyta Rabiei* selama 24 jam) dan kondisi sehat pada jaringan tanaman *Chickpea* jenis *Lasseter* muda, sebagai kasus data hasil penelitian Harijati (2007).

Hasil pengelompokan ID gen yang diperoleh pada penelitian ini berbeda dengan hasil pengelompokan ID gen yang diperoleh pada penelitian Harijati

(2007). Hal ini disebabkan karena pendekatan metode yang digunakan berbeda dimana pada penelitian Harijati menggunakan metode statistik t (non-Bayesian) dengan distribusi data unimodal. Sedangkan pada penelitian ini digunakan metode Bayesian *mixture averaging*. Selain itu, hasil validasi model dengan *goodness of fit* KS menunjukkan bahwa metode Bayesian *mixture averaging* lebih baik daripada metode statistik t (non-Bayesian) yang digunakan dalam Harijati.

Metode BMMA yang telah berhasil dibangun berdasarkan kasus data *Chickpea* di Indonesia ini dapat digunakan pada berbagai kasus data lain yang memenuhi persyaratan penggunaan metode BMMA. Oleh karena itu, metode BMMA ini tidak saja berlaku untuk kasus data *Chickpea* tetapi dapat digeneralisasikan untuk berbagai kasus data lain yang memiliki distribusi *mixture*.

**Tabel 7.7.** *Goodness of Fit* KS model BMM-Normal dan BMMA-Normal beserta Hasil Pengelompokan ID Gen

No	ID Gen	<i>Goodness of Fit</i> Kolmogorov-Smirnov (KS) dan Hasil Pengelompokan ID Gen									
		Data Simulasi					Data <i>Chickpea</i>				
		OLS	BMM-Normal		BMMA-Normal		OLS	BMM-Normal		BMMA-Normal	
Fungsi Ketahanan:			Nilai KS	Hasil Pengelompokan	Nilai KS	Hasil Pengelompokan		Nilai KS	Hasil Pengelompokan	Nilai KS	Hasil Pengelompokan
1	LS0024	<b>0,2670</b>	0,2785	<i>Regulated</i>	0,2745	<i>Regulated</i>	0,2534	0,2590	<i>Regulated</i>	<b>0,2422</b>	<i>Regulated</i>
2	LS0616	0,3479	0,3168	<i>Regulated</i>	<b>0,3159</b>	<i>Regulated</i>	0,3800	0,3325	<i>Regulated</i>	<b>0,3524</b>	<i>Regulated</i>
3	LS0159	0,3058	0,2935	<i>Regulated</i>	<b>0,2883</b>	<i>Up-regulated</i>	0,2929	<b>0,2546</b>	<i>Up-regulated</i>	0,2741	<i>Up-regulated</i>
4	LS0162	0,2793	0,2791	<i>Regulated</i>	<b>0,2673</b>	<i>Regulated</i>	0,2692	0,2707	<i>Regulated</i>	<b>0,2588</b>	<i>Regulated</i>
5	LS0185	0,2877	0,2861	<i>Regulated</i>	<b>0,2824</b>	<i>Regulated</i>	0,2752	0,2750	<i>Regulated</i>	<b>0,2573</b>	<i>Regulated</i>
6	LS0081	0,3465	0,3181	<i>Up-regulated</i>	<b>0,3167</b>	<i>Regulated</i>	0,3427	0,3195	<i>Up-regulated</i>	<b>0,3032</b>	<i>Regulated</i>
7	LS0752	0,3417	0,3163	<i>Up-regulated</i>	<b>0,3299</b>	<i>Up-regulated</i>	0,3584	0,3226	<i>Up-regulated</i>	<b>0,3211</b>	<i>Up-regulated</i>
8	LS0688	0,2637	0,2725	<i>Up-regulated</i>	<b>0,2631</b>	<i>Up-regulated</i>	0,2554	0,2525	<i>Up-regulated</i>	<b>0,2394</b>	<i>Regulated</i>
9	U273	0,3477	0,3227	<i>Up-regulated</i>	<b>0,3176</b>	<i>Up-regulated</i>	0,3415	<b>0,2985</b>	<i>Up-regulated</i>	0,3156	<i>Up-regulated</i>
10	U070	0,2344	0,2699	<i>Regulated</i>	<b>0,2179</b>	<i>Regulated</i>	0,2291	0,2742	<i>Up-regulated</i>	<b>0,2242</b>	<i>Regulated</i>

**Tabel 7.7.** (Lanjutan)

No	ID Gen	Goodness of Fit Kolmogorov-Smirnov (KS) dan Hasil Pengelompokan ID Gen									
		Data Simulasi					Data Chickpea				
		OLS	BMM-Normal		BMMA-Normal		OLS	BMM-Normal		BMMA-Normal	
Fungsi Energi:	Nilai KS		Hasil Pengelompokan	Nilai KS	Hasil Pengelompokan	Nilai KS		Hasil Pengelompokan	Nilai KS	Hasil Pengelompokan	
11	U325	0,3275	0,3161	Up-regulated	0,2997	Up-regulated	0,3147	0,2982	Up-regulated	0,2863	Up-regulated
12	U385	0,2774	0,2894	Regulated	0,2774	Regulated	0,2662	0,2550	Up-regulated	0,2466	Regulated
13	U447	0,3318	0,3183	Regulated	0,3007	Regulated	0,3178	0,3114	Regulated	0,2871	Regulated
14	CA0426	0,2986	0,2769	Regulated	0,2643	Regulated	0,2890	0,2599	Regulated	0,2471	Regulated
15	CA0534	0,2896	0,3891	Up-regulated	0,3917	Up-regulated	0,3318	0,3527	Up-regulated	0,3640	Up-regulated
16	CA0702	0,2892	0,3044	Up-regulated	0,3108	Up-regulated	0,2743	0,2746	Up-regulated	0,2715	Regulated
17	CA1120	0,3684	0,3451	Regulated	0,3319	Regulated	0,3656	0,3381	Regulated	0,3241	Regulated
18	U001	0,2304	0,2692	Regulated	0,2466	Regulated	0,2263	0,2679	Regulated	0,2349	Regulated
19	CA0810	0,2328	0,3692	Regulated	0,3233	Regulated	0,1855	0,3191	Regulated	0,2862	Regulated

## BAB VIII

### PENUTUP

Pada bab ini akan disajikan rangkuman dari hasil-hasil penelitian dan pembahasan dalam beberapa kesimpulan, saran, rekomendasi dan keberlanjutan penelitian yang merupakan elaborasi kontribusi utama dari disertasi ini.

#### 8.1. Kesimpulan

Hasil analisis data dan pembahasan yang telah dijabarkan dan disajikan dalam bab-bab sebelumnya dijadikan dasar untuk pengambilan kesimpulan berikut:

1. Pemodelan *Bayesian Mixture Model Averaging* (BMMA) yang merupakan gabungan antara pemodelan *Bayesian Mixture* (BMM) dan *Bayesian Averaging* (BMA) telah berhasil dikembangkan, baik secara teori maupun secara empiris melalui data simulasi dan diterapkan pada data *Chickpea* di Indonesia hasil penelitian Harijati (2007). Pemodelan BUM-normal, pemodelan BMA-normal unimodal, pemodelan BMM-normal, dan pemodelan BMMA-normal telah dibangun untuk data simulasi maupun data *Chickpea*. Distribusi posterior pemodelan BMMA yang telah dibangun adalah distribusi *mixture* normal dua dan tiga komponen, dimana fungsi *likelihood* dari data adalah normal dan prior dari parameter  $\lambda[j]$  berdistribusi normal, parameter  $\tau[j]$  berdistribusi gamma, dan parameter  $w[j]$  berdistribusi *categorical dirichlet*, dengan  $j = 1, 2, \dots, k$ . Seleksi model Occam's Window sebagai salah satu tahapan untuk membangun model BMMA-normal telah berhasil dibangun, dimana metode seleksi ini mampu memasukkan hampir sebagian besar model baik yang bisa dibentuk ke dalam pemodelan BMMA-normal. Pemodelan BMMA-normal yang terbangun cukup kompleks tetapi dengan kemajuan teknologi komputasi, kompleksitas model bukan menjadi masalah yang penting, lebih-lebih model BMMA-normal ini memiliki nilai *goodness of fit* KS yang paling kecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel}$  dibandingkan pemodelan yang lain, yaitu model OLS, model

BUM-normal, BMA-normal unimodal, dan BMM-normal, sehingga model BMMA-normal adalah model terbaik.

2. Beberapa algoritma telah berhasil dibangun untuk membentuk model data simulasi maupun data *Chickpea* termasuk model BMMA-normal. Ada sembilan algoritma yang telah terbangun, yaitu Algoritma 4.2 tentang Algoritma Proses Pendahuluan, Algoritma 4.3 tentang Algoritma Proses Pemulusan, Algoritma 4.4 tentang Pengembangan Algoritma RJMCMC, Algoritma 4.5 tentang Algoritma Pembangkitan Data Simulasi, Algoritma 5.1 tentang Algoritma BUM-normal, Algoritma 5.2 tentang Algoritma Seleksi Model Occam' Window, Algoritma 5.3 tentang Algoritma BMA-Normal Unimodal, Algoritma 6.1 tentang Algoritma BMM-Normal, dan Algoritma 7.1 tentang Algoritma BMMA-Normal. Semua algoritma yang telah dibangun digunakan sebagai dasar untuk membangun teknologi komputasi menggunakan Program R termasuk dengan R-WinBUGS.
3. Pemodelan BMMA-normal dua dan tiga komponen pada 19 ID gen data simulasi dan data *Chickpea*, yaitu: LS0024, LS0616, LS0159, LS0162, LS0185, LS0081, LS0752, LS0688, U273, dan U070 sebagai pembawa fungsi ketahanan, serta ID gen: U325, U385, U447, CA0426, CA0534, CA0702, CA1120, U001, dan CA0810 sebagai pembawa fungsi energi menunjukkan bahwa model ini secara umum memiliki nilai  $KS_{tabel}$  terkecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel}$  dan menjadi model terbaik dibandingkan model OLS, model BUM-normal, model BMA-normal unimodal, dan model BMM-normal. Metode seleksi model Occam's Window mampu dengan baik melibatkan hampir sebagian besar model baik yang telah dibangun, yaitu sebanyak 918 model (dari 1.000 model yang dibangkitkan) dalam pemodelan BMMA-normal. Pemodelan BMMA-normal dua dan tiga komponen pada data *Chickpea* mampu menemukan lima ID gen pada kelompok *Up-regulated*, yaitu ID Gen: LS0159, LS0752, U273, U325, dan CA0534, dimana ID gen pada kelompok ini tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei*. Lima ID gen yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* inilah yang dapat digunakan untuk menemukan varietas baru bagi tanaman *Chickpea* jenis *Lasseter* muda dengan cara penyisipan gen,

sehingga akan diperoleh jenis varietas baru dari tanaman *Chickpea* yang tangguh (*resistent*) terhadap serangan penyakit jamur patogen *Ascochyta Rabiei*. Pemodelan BMMA-normal yang telah dibangun sangat baik dalam merepresentasi data *Chickpea* dengan nilai *goodness of fit* KS yang paling kecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel}$  dibandingkan dengan model-model yang lain.

4. Pemodelan Bayesian Unimodal Normal (BUM-Normal) mampu menemukan empat ID gen pada data *Chickpea*, yaitu: LS0035, LS0612, LS0759, dan LS0341 yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei*. Empat ID gen yang tahan terhadap serangan penyakit jamur patogen *Ascochyta Rabiei* inilah yang dapat digunakan untuk menemukan varietas baru bagi tanaman *Chickpea* jenis *Lasseter* muda dengan cara penyisipan gen, sehingga akan diperoleh jenis varietas baru dari tanaman *Chickpea* yang tangguh terhadap serangan penyakit jamur patogen *Ascochyta Rabiei*. Pemodelan BUM-normal yang telah dibangun cukup baik dalam merepresentasi data *Chickpea* dengan nilai *goodness of fit* KS yang cukup kecil dan masih dalam jaminan lebih kecil dari nilai  $KS_{tabel}$ .
5. Pemodelan BUM-normal dan BMMA-normal telah menemukan sembilan ID gen (dari 26 ID yang diamati), yaitu ID gen: LS0035, LS0612, LS0759, LS0341, LS0159, LS0752, U273, U325, dan CA0534 yang tangguh terhadap serangan penyakit jamur patogen *Ascochyta Rabiei*, sehingga melalui sembilan ID gen inilah akan dikembangkan penelitian lanjut untuk menemukan jenis varietas baru yang tangguh menghadapi serangan penyakit dari tanaman *Chickpea* jenis *Lasseter* muda dengan cara melakukan penyisipan gen.

## 8.2. Saran, Rekomendasi, dan Keberlanjutan Penelitian

Berdasarkan kesimpulan yang telah diperoleh, maka beberapa saran, rekomendasi, dan keberlanjutan penelitian dapat disampaikan sebagai berikut:

1. Pemodelan BMMA yang telah dikembangkan untuk data *microarray* ini, yaitu data perbedaan ekspresi gen jaringan tanaman *Chickpea* di Indonesia masih terbatas pada jumlah komponen *mixture* sebanyak dua dan tiga komponen saja. Hal ini karena ukuran sampel data *Chickpea* hanya sebanyak  $n = 3$ . Oleh



karena itu bisa dikembangkan untuk kasus lain yang memiliki ukuran sampel  $n$  lebih besar dari tiga. Penggunaan metode Pengembangan Algoritma RJMCMC yang telah dikembangkan sebagai metode untuk menentukan jumlah komponen *mixture* dalam data akan sangat membantu untuk memilah dan mengidentifikasi jumlah komponen *mixture* yang optimal. Hal ini dimaksudkan untuk melihat sejauhmana kemampuan dari metode Pengembangan Algoritma RJMCMC yang telah dibangun.

2. Pendekatan distribusi normal telah digunakan dalam kasus data Chickpea di Indonesia dan tidak menutup kemungkinan dikembangkan dengan pendekatan distribusi data yang lain, apalagi dalam kasus data *Chickpea* ini ukuran sampel  $n$  yang mampu diamati/diperoleh hanya sebanyak  $n = 3$  karena keterbatasan kemampuan untuk mengidentifikasi data *microarray* (biaya mahal, ketersediaan data sangat terbatas, belum ada alat pendeteksi ekspresi gen di Indonesia).
3. Metode seleksi model Occam's Window telah digunakan dalam penelitian ini dan dapat dikembangkan dengan metode seleksi model yang lain, seperti metode MC<sup>3</sup> (*Markov Chain Monte Carlo Model Composition*).
4. Algoritma komputasi pemodelan BUM-normal, BMA-normal unimodal, BMM-normal, dan BMMA-normal untuk kasus data *microarray* di Indonesia menggunakan Program R telah dibangun dan dapat dimanfaatkan serta dikembangkan untuk pemodelan kasus lain yang memiliki karakteristik data yang sejenis. Algoritma komputasi ini juga dapat mengkayakan area Program R yang telah ada, khususnya untuk pemodelan kasus data *microarray* di Indonesia dengan cara diinputkan dalam *Library Software R*. Untuk kepentingan ini, algoritma komputasi yang telah dibangun masih membutuhkan penyempurnaan-penyempurnaan dalam hal manajemen dan efisiensi program sehingga lebih *user friendly*.
5. Hasil yang telah diperoleh dalam penelitian ini, khususnya informasi tentang kelompok ID gen yang berada pada area *Up-regulated* sangat bermanfaat dan dibutuhkan di bidang Biologi di Indonesia khususnya untuk Balai Penelitian Tanaman (BPT) untuk mengembangkan jenis-jenis varietas baru melalui penyisipan gen yang memiliki karakter *Up-regulated* dari suatu tanaman yang

sangat dibutuhkan oleh manusia, sehingga peningkatan produksi dari tanaman jenis varietas baru tersebut dapat berhasil dilakukan. Selain itu, melalui penelitian ini juga dapat diketahui berbagai ID gen yang sangat rentan terhadap serangan penyakit, sehingga para peneliti di bidang Biologi mampu mengembangkan sebuah metode untuk mengatasinya.

*Halaman ini sengaja dikosongkan*

## KARYA-KARYA YANG TELAH DIPUBLIKASIKAN DI JURNAL INTERNASIONAL

**Tabel 7.8.** Karya di Jurnal dan *Proceeding* Internasional

No	Jurnal	Nama Jurnal/Proceeding	Judul	Tanggal/Tahun	Lokasi/Alamat URL*
1	Jurnal Internasional (Scopus)  Astuti dkk. (2014b)	Applied Mathematical Sciences (Volume 8, No. 146, pp. 7277-7287, 2014), DOI: <a href="http://dx.doi.org/10.12988/ams.2014.49760">http://dx.doi.org/10.12988/ams.2014.49760</a>	Kolmogorov-Smirnov and Continuous Ranked Probability Score Validation on the Bayesian Model Averaging for Microarray Data	22 Oktober 2014	<a href="http://www.m-hikari.com/ams/ams-2014/ams-145-148-2014/astutiAMS145-148-2014.pdf">http://www.m-hikari.com/ams/ams-2014/ams-145-148-2014/astutiAMS145-148-2014.pdf</a>
2	Jurnal Internasional (Scopus)	Global Journal of Pure and Applied Mathematics (Volume 10, No. 4, pp. 591-600, 2014), ISSN: 0973-1768)	Cox Regression Model for Lifetime of a Newborn Baby (Neonates) in DR. Saiful Anwar Malang Hospital	20 Nopember 2014	<a href="http://www.ripublication.com/Volumegjpamv10n4.htm">http://www.ripublication.com/Volumegjpamv10n4.htm</a>
3	Jurnal Internasional (Scopus)  Astuti dkk. (2015b)	International Journal of Applied Mathematics and Statistics (Volume 53, Issue No. 4, pp. 166-173, 2015), ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)	Occam's Window Selection in Bayesian Model Averaging Modeling for Gene Expression Data from Chickpea Plant	8 Juni 2015	<a href="http://www.ceser.in/ceserp/index.php/ijamas/issue/view/295">http://www.ceser.in/ceserp/index.php/ijamas/issue/view/295</a>

\*Fullpaper dapat diakses di alamat URL jurnal

**Tabel 7.8. (Lanjutan)**

No	Jurnal	Nama Jurnal/Proceeding	Judul	Tanggal/Tahun	Lokasi/Alamat URL*
4	Jurnal Internasional (Scopus)  Astuti dkk. (2015c)	Journal of Mathematics and Statistics (Volume 11, No. 2, pp. 45-51, 2015), DOI: 10.3844/jmssp.2015.45.51	An Algorithm for Determining the Number of Mixture Components on the Bayesian Mixture Model Averaging for Microarray Data	28 September 2015	<a href="http://thescipub.com/issue-jmss/11/2">http://thescipub.com/issue-jmss/11/2</a>
5	Jurnal Internasional (Scopus)  Astuti dkk. (2016a)	Global Journal of Pure and Applied Mathematics (Volume 12, No. 4, pp. 3143-3158, 2016), ISSN: 0973-1768)	Blood Sugar Levels of Diabetes Mellitus Patients Modeling with Bayesian Mixture Model Averaging	8 Agustus 2016	<a href="http://www.ripublication.com/gjpam/16/gjpamv12n4_32.pdf">http://www.ripublication.com/gjpam/16/gjpamv12n4_32.pdf</a>
6	Jurnal Internasional (Scopus)  Astuti dkk. (2016c)	Communications in Statistics, DOI: <a href="http://dx.doi.org/10.1080/03610926.2016.1239112">http://dx.doi.org/10.1080/03610926.2016.1239112</a>	Bayesian Mixture Model Averaging (BMMA) for Identifying the Differences Genes Expression of Chickpea ( <i>Cicer Arietinum</i> ) Plant Tissue	14 Oktober 2016	<a href="http://www.tandfonline.com/doi/full/10.1080/03610926.2016.1239112">http://www.tandfonline.com/doi/full/10.1080/03610926.2016.1239112</a>

\*Fullpaper dapat diakses di alamat URL jurnal

**Tabel 7.8. (Lanjutan)**

No	Jurnal	Nama Jurnal/Proceeding	Judul	Tanggal/Tahun	Lokasi/Alamat URL*
7	Jurnal Internasional (Non Scopus)  Astuti dkk. (2014a)	Natural-A Journal of Scientific Modeling and Computation (Volume 1, No. 2, pp. 67-74, 2014), ISSN: 2303-0135	Model Components Selection in Bayesian Model Averaging Using Occam's Window for Microarray Data	2 Juni 2014	<a href="http://natural-a.ub.ac.id/index.php/natural-a/issue/view/14/showToc">http://natural-a.ub.ac.id/index.php/natural-a/issue/view/14/showToc</a>
8	Proceeding Internasional (Scopus)  Astuti dkk. (2016d)	Proceeding Scopus IEEE-Xplore (Accepted)	Bayesian Mixture Modeling for Blood Sugar Levels of Diabetes Mellitus Patients (Case Study in RSUD Saiful Anwar Malang Indonesia)	1 Nopember 2016	<a href="#">Diperkirakan publish online September 2017</a>

\*Fullpaper dapat diakses di alamat URL jurnal

*Halaman ini sengaja dikosongkan*

## DAFTAR PUSTAKA

- Aitkin, M. (2001). Likelihood and Bayesian Analysis of Mixtures. *Statistical Modelling*. 1: 287-304.
- Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. dan Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *In Proceedings of the National Academy of Sciences*. Vol. 96: 6745–6750.
- Antoniak, C. E. (1974). Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics*. 2: 1152–1174.
- Astuti, E. Y. (2006). Implementasi Bayesian Markov Chain Monte Carlo pada Pemodelan Portopolio Optimal dengan Pendekatan Model *Mixture* dari Beberapa *Mixture*. *Tesis*. Pascasarjana. ITS, Surabaya.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2012). Theoretical Studies of Bayesian Mixture Model Averaging Methods for Microarray Data. *Proceeding of 2<sup>nd</sup> Regional Conference on Applied and Engineering Mathematics (RCAEM-II)*. Penang, Malaysia. ISBN 978-967-11189-00.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2013). Highest Posterior Density for Identifying Differences in Gene Expression Microarray Experiments. *Proceeding of the 3<sup>rd</sup> Annual Basic Science International Conference (BaSIC 2013)*. Volume 3. UB, Malang. ISSN 2338-0136.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2014a). Model Components Selection in Bayesian Model Averaging Using Occam's Window for Microarray Data. *Natural-A Journal of Scientific Modeling and Computation*, 2 (1): 67-74. ISSN: 2303-0135.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2014b). Kolmogorov-Smirnov and Continuous Ranked Probability Score Validation on the Bayesian Model Averaging for Microarray Data. *Applied Mathematical Sciences*, 146 (8): 7277-7287. DOI: <http://dx.doi.org/10.12988/ams.2014.49760>.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015a). Bayesian Model Averaging dengan Metode Seleksi Komponen Model Occam's Window pada Data Diabetes Mellitus. Seminar Nasional SEMIRATA Bidang MIPA 2015. Universitas Tanjungpura, Pontianak.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015b). Occam's Window Selection in Bayesian Model Averaging Modeling for Gene Expression Data from *Chickpea* Plant. *International Journal of*



- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015c). An Algorithm for Determining the Number of Mixture Components on the Bayesian Mixture Model Averaging for Microarray Data. *Journal of Mathematics and Statistics*, 11 (2): 45-51. DOI: 10.3844/jmssp.2015.45.51.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015d). Bayesian Model Averaging Modeling for Blood Sugar Levels of Diabetes Mellitus Patients in RSUD Saiful Anwar Malang. *International Symposium on Biomathematics (Symomath)* 2015. ITB, Bandung.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016a). Blood Sugar Levels of Diabetes Mellitus Patients Modeling with Bayesian Mixture Model Averaging. *Global Journal of Pure and Applied Mathematics*, 12 (4): 3143–3158. ISSN 0973-1768.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016b). Goodness of Fits Model with Kolmogorov-Smirnov on the Bayesian Mixture Model Averaging (Case Study on Diabetes Mellitus Patients at RSUD Saiful Anwar Malang). *International Symposium on Biomathematics (Symomath)* 2016. Universitas Hassanudin, Makassar.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016c). Bayesian Mixture Model Averaging (BMMA) for Identifying the Differences Genes Expression of Chickpea (*Cicer Arietinum*) Plant Tissue. *Communications in Statistics-Theory and Methods*, Publish Online Accepted Manuscript Version. DOI: <http://dx.doi.org/10.1080/03610926.2016.1239112>.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016d). Bayesian Mixture Modeling for Blood Sugar Levels of Diabetes Mellitus Patients (Case Study in RSUD Saiful Anwar Malang Indonesia). *Accepted publish on Proceeding of IEEEExplore* 2017.
- Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016e). Seleksi Model Occam's Window dalam Pemodelan Bayesian Model Averaging (Studi Kasus Diabetes Mellitus di RSUD Saiful Anwar Malang). Konferensi Nasional Matematika (KNM) ke-18 2016. Universitas Riau, Pekanbaru.
- Baggerly, K. A., Coombes, K. R., Hess, K. R., Stivers, D. N., Abruzzo, L. V. dan W., Z. (2001). Identifying Differentially Expressed Genes in cDNA Microarray Experiments. *Journal Computational Biology*. 8: 639–659.

- Baldi, P. dan Long, A. D. (2001). A Bayesian Framework for the Analysis of Microarray Expression Data: Regularized T-Test dan Statistical Inferences of Gene Changes. *Bioinformatics*. 17: 509-519.
- Box, G. E. P. dan Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. MA: Addison-Wesley, Massachusetts.
- Cassela, G. dan George, I. E. (1992). Explaining the Gibbs Sampler. *Journal of The American Statistician Association*. 46(3): 167-174.
- Charlin, B. P. dan Chip, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of Royal Statistical Society. B*. 57(3): 473-484.
- Chen, Y., Dougherty, E., dan Bittner, M. (1997). Ratio-based Decisions and the Quantitative Analysis of cDNA Microarray Images. *Journal of Biomedical Optics*. 2: 364-374.
- Congdon, P. (2006). *Bayesian Statistical Modelling* (2 ed.). USA: John Wiley & Sons.
- Cryer, J. D. dan Chan, K. S. (2008). *Time Series Analysis With Application in R*. Springer, New York.
- Darlington, R. B. dan Hayes, A. F. (2017). *Regression Analysis and Linear Models, Concepts, Applications and Implementation*. The Guilford press, New York, London.
- DeRisi, J. L., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., dan Trent, J. M. (1996). Use of a cDNA Microarray to Analyze Gene Expression Patterns in Human Cancer. *Nature Genetics*. 14: 457-460.
- Dinar, Y. K. L. (2009). Estimasi Densitas Pengeluaran Rumah Tangga Kota Mataram Hasil Survei Biaya Hidup 2007 dengan Pendekatan Bayesian Mixture of Mixture. *Tesis*. Pascasarjana. ITS, Surabaya.
- Do, K. A., Muller P., dan Tang, F. (2004). A Bayesian Mixture Model For Differential Gene Expression. *Research in Department of Biostatistics*, The University of Texas.
- Dudoit, S., Yang, Y. H., Callow, M. J., dan Speed, T. P. (2001). Statistical Methods for Identifying Genes with Differential Expression in Replicated cDNA Microarrays Experiments". *Statistica Sinica*. 12: 111-139.

- Duggan, J. D., Bittner, M., Chen, Y., Meltzer, P. dan Trent, J. M. (1999). Expression Profiling Using CDNA Microarrays. *Nature Genetics*. 21: 10-14.
- Efron, B., Tibshirani, R., Storey, J. D., dan Tusher, V. (2001). Empirical Bayes Analysis of a Microarray Experiment. *Journal of the American Statistical Association*. 96: 1151–1160.
- Escobar, M. D. dan West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*. 90: 577–588.
- Feng, D. (2013). Miscellaneous Functions. Package ‘miscF’. Version 0.1-0. License GPL.
- Gamerman, D. (1997). *Markov Chain Monte Carlo*. Chapman & Hall, London.
- Gelman, A., Carlin, J. B., Stern, H. S. dan Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.
- Giles, P. J. dan Kipling, D. (2003). Normality of Oligonucleotide Microarray Data and Implications for Parametric Statistical Analyses. *Bioinformatics*. 19: 2254-2262.
- Gilks, W. R. (1995). *Full Conditional Distribution*. In Markov Chain Monte Carlo in Practise. (eds). Gilks, W. R., Richardson, S. dan Spiegelhalter, D. J. Chapman & Hall, London. 75-88.
- Gosh, J. K., Delampady, M. dan Samanta, T. (2006). *An Introduction to Bayesian Analysis Theory and Methods*. Springer, New York.
- Govindarajan R., Duraiyan, J., Kaliyappan K. dan Palanisamy, M. (2012). Microarray and its Applications. *Journal of Pharmacy and Bioallied Sciences*. 4: S310-S312. DOI: [10.4103/0975-7406.100283](https://doi.org/10.4103/0975-7406.100283).
- Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*. 82 (4). 711-732.
- Harijati, N. (2007). A Study of the Resistance of *Chickpea (Cicer Arietinum)* to *Ascochyta Rabiei* and the Effect of Age of Plant Tissue on Disease Development. *Ph.D. Thesis*. La Trobe University, Australia.
- Hastings, W. K. (1970). Monte Carlo Sampling Methods using Markov Chain and their Applications. *Biometrika*. 57: 97-109.

- Hoeting, J. A., Madigan, D., Raftery, A. E., dan Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial (with Discussion). *Statistical Science*.14: 382-417.
- Hustianda, V. F. (2012). Perbandingan Bayesian Model Averaging dan Regresi Linear Berganda dalam Memprediksi Faktor-Faktor yang Mempengaruhi Jumlah Kematian Bayi di Jawa Timur. *Tesis*. Jurusan Statistika. FMIPA-ITS, Surabaya.
- Ibrahim, J. G., Chen, M. H., dan Gray, R. J. (2002). Bayesian Models for Gene Expression With DNA Microarray Data. *Journal of the American Statistical Association*. 97: 88-99.
- Iriawan, N. (2000). Computationally Intensive Approaches to Inference in Neo-Normal Linear Models. *Ph.D. Thesis*. CUT-Australia.
- Iriawan, N. (2001). Penaksiran Model Mixture Normal Univariabel: Suatu Pendekatan Metode Bayesian dengan MCMC. *Prosiding Seminar Nasional dan Konferda VII Matematika Wilayah DIY & Jawa Tengah*. Yogyakarta. 105-110.
- Iriawan, N. (2003a). *Teknik Simulasi*. Modul Ajar. ITS, Surabaya.
- Iriawan, N. (2003b). *Pemodelan Data dengan MCMC Menggunakan WinBUGS 1.4*. Modul Ajar. ITS, Surabaya.
- Iriawan, N. (2012). *Pemodelan dan Analisis Data-Driven*. Volume I. ITS Press, Surabaya.
- Jeffreys, H. (1961). *Theory of Probability*. 3<sup>rd</sup> ed. Oxford Classic Texts in the Physical Sciences. Oxford University. Press, Oxford.
- Kass, R. dan Raftery, A. (1995). Bayes Factors and Model Uncertainty. *Journal of the American Statistical Association*. 90: 773-795.
- Knudsen, S. (2004). *A Guide to Analysis of DNA Microarray Data*. Second Edition. John Wiley & Sons, Inc., New Jersey, Canada.
- Korenberg, M. J. (2007). *Microarray Data Analysis, Methods and Applications*. Humana Press, Totowa, New Jersey.
- Kuswanto, H. dan Sari, M. R. (2013). Bayesian Model Averaging with Markov Chain Monte Carlo for Calibrating Temperature Forecast from Combination of Time Series Models. *Journal of Mathematics and Statistics*, 9(4): pp. 349.
- Law, A. M. dan Kelton, W. D. (2000). *Simulation Modeling and Analysis*. 3<sup>rd</sup> Edition. McGraw Hill, Singapore.

- Liang, F. M, Troung, Y, dan Wong, W. H. (2001). Automatic Bayesian Model Averaging for Linear Regression and Applications in Bayesian Curve Fitting. *Statistical Science*, 11(4): 1005-1029.
- Lonnstedt, I. dan Speed, T. P. (2002). Replicated Microarray Data. *Statistica Sinica*. 12: 31–46.
- Madigan, D. dan Raftery, A. E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *Journal of the American Statistical Association*. Vol.89. 428: 1535-1546.
- Maindonald, J. H. (2008). *Using R for Data Analysis and Graphics Introduction, Code and Commentary*. Centre for Mathematics and Its Applications, Australian National University.
- Marin, J. M., Mengersen, K. dan Robert, C. P. (2005). Bayesian Modeling and Inference on Mixtures of Distribution. *Handbook of Statistics*. Vol. 25: 50.
- McLachlan, G. J. dan Basford, K. E. (1988). *Mixture Models Inference and Applications to Clustering*. Marcel Dekker, New York.
- McLachlan, G. J., Bean, R. W. dan Peel, D. (2002). A mixture Model-Based Approach to the Clustering of Microarray Expression Data. *Bioinformatics*. 18: 413–422.
- Mengersen, K. (2009). Modul 1 Bayesian Analysis. *Short Course on Bayesian Modeling*. Jurusan Statistika. ITS, Surabaya.
- Montgomery, J. dan Nyhan, B. (2010). Bayesian Model Averaging: Theoretical Developments and Practical Applications. *Society for Political Methodology working paper*.
- Muller, P., Parmigiani, G., Robert, C., dan Rousseau, J. (2002). Optimal Sample Size for Multiple Testing: the Case of Gene Expression Microarrays. *Tech. rep.*, University of Texas, M.D. Anderson Cancer Center.
- Nadon, R. dan Shoemaker, J. (2002). Statistical Issues with Microarrays: Processing and Analysis". *Trends in Genetics*. 18: 265-271.
- Newton, M. A., Kendzierski, C. M., Richmond, C. S., Blattner, F. R., dan Tsui, K. W. (2001). On The Differential Variability of Expression Ratios: Improving Statistical Inference About Gene Expression Changes from Microarray Data". *Journal of Computational Biology*. 8(1): 37-52.

- Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. New Jersey: John Wiley & Sons, Inc.
- Pan, W., Lin, J. dan Le, C. T. (2002). Model-Based Cluster Analysis of Microarray Gene Expression Data. *Genome Biology*. 3(2): Research009.1–Research009.9.
- Prihartanti, W. (2002). Studi Realibilitas Model Kerusakan Ganda: Pendekatan *Mixture Weibull*. Tesis. Pascasarjana. ITS, Surabaya.
- Purnamasari, R. (2011). Penggunaan Metode Bayesian Model Averaging (BMA) dengan Pendekatan Markov Chain Monte Carlo (MCMC) untuk Peramalan Kecepatan Angin Rata-Rata Harian Stasiun Meteorologi Juanda. *Tugas Akhir*. Jurusan Statistika. FMIPA-ITS, Surabaya.
- Raftery, A. E, Madigan, D, dan Hoeting, J. (1997). Bayesian Model Averaging for Linear Regression Models. *Journal of the American Statistical Association*. 92.
- Richardson, S. dan Green, P. J. (1997). On Bayesian Analysis with an Unknown Number of Components. *Journal of the Royal Statistical Society*. B. 59. 4: 731–792.
- Robert, C. P. (1996). *Mixture of Distributions: Inference and Estimation*. In Markov Chain Monte Carlo in Practice. eds. W. R. Gilks, S. R. dan Spiegelhalter, D. J. London: Chapman & Hall. 441–464.
- Schena, M., Shalon, D., Davis, R. W. dan Brown, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*. 270: 467–470.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. dan Davis, R. W. (1996). Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences*. 93: 10614–10619.
- Sebastiani, P., Gussoni, E., Kohane, I. S., dan Ramoni, M. (2003). Statistical Challenges in Functional Genomics (with Discussion). *Statistical Science*. 18: 33-70.
- Sebastiani, P., Xie H., dan Ramoni, M.F. (2006). Bayesian Analysis Of Comparative Microarray Experiments By Model Averaging. *International Society For Bayesian Analysis*. 1. 4: 707-732.
- Sheskin, D. J. 2007. *Handbook of Parametric and Nonparametric Statistical Procedures*, 4<sup>th</sup> edition. Chapman & Hall/CRC.

- Shoemaker, J. S. dan Lin, S. M. (2005). *Methods of Microarray Analysis IV*. Springer, New York.
- Stafford, P. (2008). *Methods in Microarray Normalization*. CRC Press., New York.
- Stephens, M. (1997). Bayesian Methods for Mixtures of Normal Distribution. *Ph.D. Thesis*. Oxford.
- Suharjo, B. (2003). Implementasi Bayesian Markov Chain Monte Carlo pada Distribusi Mixture Normal. *Tesis*. Pascasarjana. ITS, Surabaya.
- Suhartono. (2009). *Analisis Data Statistik dengan R*. Graha Ilmu, Yogyakarta.
- Tanner, M. A. (1996). *Tools for Statistical Inference: Methods to the Exploration of Posterior Distributions and Likelihood Functions*. 3<sup>th</sup> ed. Springer-Verlag, New York.
- Thomas, A., Spiegelhalter, D. J., dan Gilks, W. R. (1992). *BUGS: A Program to Perform Bayesian Inference using Gibbs Sampling*. In Bernardo, J., Berger, J.
- Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*. 22: 1701–1762.
- Thomas, J. G., Olson, J. M., Tapscott, S. J. dan Zhao, L. P. (2001). An Efficient and Robust Statistical Modeling Approach to Discover Differentially Expressed Genes Using Genomic Expression Profiles. *Genome Research*. 11: 1227-1236.
- Tusher, V. G., Tibshirani, R. dan Chu G. (2001). Significance Analysis of Microarrays Applied to the Ionizing Radiation Response. *In Proceedings of the National Academy of Sciences*. Vol. 98: 5116–5121.
- Viallefont, V., Raftery, A.E. dan Richardson, S. (2001). Variable Selection and Bayesian Model Averaging in Case- Control Studies. *Statistical in Medicine*.
- Volinsky, C. T. (1997). Bayesian Model Averaging for Censored Survival Models. *Ph.D. Dissertation*, Univ. Washington, Seattle.
- Walpole, R. E. (1993). *Pengantar Statistika*. PT Gramedia Pustaka Utama, Jakarta.
- Walsh, B. (2004). *Markov Chain Monte Carlo and Gibbs Sampling*. Lecture Notes for EEB 581, version 26.

- Wati, D. M. (2006). Analisis Bayesian Markov Chain Monte Carlo pada Pemodelan Mixture Normal dengan Banyak Komponen Tidak Diketahui (Studi Kasus: Tingkat Inflasi di Indonesia). *Tesis*. Pascasarjana. ITS, Surabaya.
- Wit, E. dan McClure, J. (2004). *Statistics for Microarrays. Design, Analysis, and Inference*. John Wiley & Sons Ltd., England.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. John Wiley, New York.



*Halaman ini sengaja dikosongkan*

## LAMPIRAN

### Lampiran 1. Code Program R-WinBUGS untuk Membangkitkan Data Simulasi, Membangun Model OLS, BUM-Normal, BMA-Normal, BMM-Normal, dan BMMA-Normal

```
# ketik proc_gendata(y,tol), dimana:
#      y      :      data asli (hasil pengamatan)
#      tol      :      toleransi selisih terhadap data asli (default 0.0999)
#
#Output
#      data hasil bangkitan
##-----

      y <- sort(y)
      n <- length(y)
      miu <- mean(y)
      std <- sd(y)
      yRnd <- 0
      for(i in 1:n){
        if(i==1|i==n){
          repeat{
            temp <- rnorm(1,miu,std)
            evals <- abs(y[i]-temp)
            if(temp>=y[1]&temp<=y[n]&evals<tol){
              break
            }
          }
        } else {
          repeat{
            temp <- rnorm(1,miu,std)
            evals <- abs(y[i]-temp)
            if(evals<tol){
              break
            }
          }
        }
        yRnd[i] <- temp
      }
      return(yRnd)
}
```

#### Run,R:

```
#Preparataion
library(BayesNormAv)
library(xlsx)
library(R2OpenBUGS)
library(nleqslv)
library(ggplot2)
idgen <- c('CA0426','CA0534','CA0702','CA0810','CA1120','LS0024',
           'LS0035','LS0081','LS0132','LS0159','LS0162','LS0185',
           'LS0341','LS0612','LS0616','LS0688','LS0752','LS0759',
           'U001','U017','U070','U273','U278','U325','U385',
           'U447')
nsim <- 1000 #number of simulation
ekse <- c(1,2) #idgen index to process
wd <- "D:Resim/LY-24" #directory
setwd(wd)
source('gen1k.r')
source('gen2k.r')
source('gen3k.r')

run <- function(y,name='raw',plot=TRUE,bugssamp=5000,append=FALSE){
#-----
#Normal pdf modeling by simultan process
#
#input
```

## Lampiran 1. (Lanjutan)

```
#      y      :      data
#      name    :
#      plot    :      plot optiona (TRUE/FALSE)
#      bugsamp :      number of bugs sample
#      append  :      write append option (TRUE/FALSE)
#output
#      output process written in txt format
#note
#      run with #Simulation part code (source this file to R)
#-----
cat('Pemrosesan BUGS',name,'...\n')

#convetional model
cat('\tNormal unimodal konvensional...\n')
mod0 <- proc_model0(y)
fcomp0 <- proc_component(mod0)
par0 <- c(1,fcomp0)
names(par0) <- c('w.1','lambda.1','sigma.1')
write.table(rbind(par0),'um/um_mean.txt',append=append,col.names=!append,row.names=FALSE)
if(plot==TRUE){
  proc_plotmodel(y=y,model=mod0,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
  ggsave(filename=paste0('um/',name,'_um.wmf'),width=14.07,height=12.85,units='cm')
}
ks0 <- proc_ksnorm(y,mod0)

##bugs estimation 1k
cat('\tBayesian normal 1k...\n')
cat('\tEstimasi...\n')
bug1 <- proc_1k(y)
simpar1 <- proc_takebugs(bug1)
write.table(simpar1,paste0('1k/',name,'.txt'))

#bum
cat('\tbum...\n')
par1 <- colMeans(simpar1)
write.table(rbind(par1),'bum/bum_mean.txt',append=append,col.names=!append,row.names=FALSE)
mod1 <- proc_bmodel(simpar1)
fcomp1 <- proc_component(mod1)
write.table(rbind(fcomp1),'bum/bum_component.txt',append=append,col.names=!append,row.names=FALSE)
if(plot==TRUE){
  proc_plotmodel(y=y,model=mod1,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
  ggsave(filename=paste0('bum/',name,'_bum.wmf'),width=14.07,height=12.85,units='cm')
}
cat('\t\tks...\n')
ks1 <- proc_ksnorm(y,mod1)
cat('\t\tthpd...\n')
hpd1 <- proc_hpd(y,mod1)
write.table(rbind(hpd1),'bum/bum_hpd.txt',append=append,col.names=!append,row.names=FALSE)

#bma
cat('\tbma...\n')
cat('\t\toccams window...\n')
occam1 <- proc_occamswindow(simpar1)
write.table(occam1,paste0('bma/occams window/',name,'.txt'))
accept1 <- nrow(occam1)
reject1 <- bugssamp-accept1
agg1 <- c(accept1,reject1,colMeans(occam1));names(agg1)[c(1,2)] <- c('naccept','nreject')
write.table(rbind(agg1),paste0('bma/bma_mean.txt'),append=append,col.names=!append,row.names=FALSE)
mod1a <- proc_bAvmodel(occam1)
fcomp1a <- proc_component(mod1a)
write.table(rbind(fcomp1a),'bma/bma_component.txt',append=append,col.names=!append,row.names=FALSE)
if(plot==TRUE){
  proc_plotmodel(y=y,model=mod1a,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
  ggsave(filename=paste0('bma/',name,'_bma.wmf'),width=14.07,height=12.85,units='cm')
}
cat('\t\tks...\n')
ks1a <- proc_ksnorm(y,mod1a)
cat('\t\tthpd...\n')
hpd1a <- proc_hpd(y,mod1a)
```

## Lampiran 1. (Lanjutan)

```

write.table(rbind(hpd1a), 'bma/bma_hpd.txt', append=append, col.names=!append, row.names=FALSE)

##bugs estimation 2k
cat("\t\tBayesian normal 2k...\n")
cat("\t\tEstimasi...\n")
bug2 <- proc_2k(y)
simpar2 <- proc_takebugs(bug2)
write.table(simpar2, paste0('2k/', name, '.txt'))

#bmm2
cat("\t\tbmm2...\n")
par2 <- colMeans(simpar2)
write.table(rbind(par2), 'bmm2/bmm2_mean.txt', append=append, col.names=!append, row.names=FALSE)
mod2 <- proc_bmodel(simpar2)
fcomp2 <- proc_component(mod2)
write.table(rbind(fcomp2), 'bmm2/bmm2_component.txt', append=append, col.names=!append,
row.names=FALSE)
if(plot==TRUE){
  proc_plotmodel(y=y, model=mod2, main="", xlab='Perbedaan Ekspresi Gen', ylab='Densitas')
  ggsave(filename=paste0('bmm2/', name, '_bmm2.wmf'), width=14.07, height=12.85, units='cm')
}
cat("\t\t\tks...\n")
ks2 <- proc_ksnorm(y, mod2)
cat("\t\t\tthpd...\n")
hpd2 <- proc_hpd(y, mod2)
write.table(rbind(hpd2), 'bmm2/bmm2_hpd.txt', append=append, col.names=!append, row.names=FALSE)

#bmma2
cat("\t\tbmma2...\n")
cat("\t\t\toccams window...\n")
occam2 <- proc_occamswindow(simpar2)
write.table(occam2, paste0('bmma2/occams window/', name, '.txt'))
accept2 <- nrow(occam2)
reject2 <- bugssamp-accept2
agg2 <- c(accept2, reject2, colMeans(occam2)); names(agg2)[c(1,2)] <- c('naccept', 'nreject')

write.table(rbind(agg2), paste0('bmma2/bmma2_mean.txt'), append=append, col.names=!append, row.names=FALSE)

SE) mod2a <- proc_bAvmodel(occam2)
fcomp2a <- proc_component(mod2a)
write.table(rbind(fcomp2a), 'bmma2/bmma2_component.txt', append=append, col.names=!append, row.names=FALSE)

LSE) if(plot==TRUE){
  proc_plotmodel(y=y, model=mod2a, main="", xlab='Perbedaan Ekspresi Gen', ylab='Densitas')
  ggsave(filename=paste0('bmma2/', name, '_bmma2.wmf'), width=14.07, height=12.85, units='cm')
}
cat("\t\t\tks...\n")
ks2a <- proc_ksnorm(y, mod2a)
cat("\t\t\tthpd...\n")
hpd2a <- proc_hpd(y, mod2a)
write.table(rbind(hpd2a), 'bmma2/bmma2_hpd.txt', append=append, col.names=!append, row.names=FALSE)
##bugs estimation 3k
cat("\t\tBayesian normal 3k...\n")
cat("\t\tEstimasi...\n")
bug3 <- proc_3k(y)
simpar3 <- proc_takebugs(bug3)
write.table(simpar3, paste0('3k/', name, '.txt'))

#bmm3
cat("\t\tbmm3...\n")
par3 <- colMeans(simpar3)
write.table(rbind(par3), 'bmm3/bmm3_mean.txt', append=append, col.names=!append, row.names=FALSE)
mod3 <- proc_bmodel(simpar3)
fcomp3 <- proc_component(mod3)
write.table(rbind(fcomp3), 'bmm3/bmm3_component.txt', append=append, col.names=!append,
row.names=FALSE)
if(plot==TRUE){
  proc_plotmodel(y=y, model=mod3, main="", xlab='Perbedaan Ekspresi Gen', ylab='Densitas')

```

## Lampiran 1. (Lanjutan)

```

}
cat("\t\t\tks...\n")
ks3 <- proc_ksnorm(y,mod3)
cat("\t\t\tthpd...\n")
hpd3 <- proc_hpd(y,mod3)
write.table(rbind(hpd3),'bmm3/bmm3_hpd.txt',append=append,col.names=!append,row.names=FALSE)

#bmma3
cat("\t\t\tbmma3...\n")
cat("\t\t\ttoccams window...\n")
occam3 <- proc_occamswindow(simpar3)
write.table(occam3,paste0('bmma3/occams window/',name,'.txt'))
accept3 <- nrow(occam3)
reject3 <- bugssamp-accept3
agg3 <- c(accept3,reject3,colMeans(occam3));names(agg3)[c(1,2)] <- c('naccept','nreject')
write.table(rbind(agg3),paste0('bmma3/bmma3_mean.txt'),append=append,col.names=!append,
row.names=FALSE)
mod3a <- proc_bAvmodel(occam3)
fcomp3a <- proc_component(mod3a)
write.table(rbind(fcomp3a),'bmma3/bmma3_component.txt',append=append,col.names=!append,
row.names=FALSE)
if(plot==TRUE){
  proc_plotmodel(y=y,model=mod3a,main=" ,xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
  ggsave(filename=paste0('bmma3/',name,'_bmma3.wmf'),width=14.07,height=12.85,units='cm')
}
cat("\t\t\tks...\n")
ks3a <- proc_ksnorm(y,mod3a)
cat("\t\t\tthpd...\n")
hpd3a <- proc_hpd(y,mod3a)
write.table(rbind(hpd3a),'bmma3/bmma3_hpd.txt',append=append,col.names=!append,row.names=FALSE)

#save ks
ks <- c(ks0,ks1,ks1a,ks2,ks2a,ks3,ks3a)
names(ks) <- c('um.ks','um.dec','bum.ks','bum.dec','bma.ks','bma.dec','bmm2.ks','bmm2.dec',

'bmma2.ks','bmma2.dec','bmm3.ks','bmm3.dec','bmma3.ks','bmma3.dec')

write.table(rbind(ks),'ks.txt',append=append,col.names=!append,row.names=FALSE)
cat('Selesai.\n\n')
}
#Simulation
for(i in ekse){
  awal <- Sys.time()

  tmpid <- idgen[i]
  setwd(paste(wd,tmpid,sep='/'))
  asldata <- scan(paste0(tmpid,'.txt'))
  gendata <- as.matrix(read.xlsx(paste0(tmpid,'-00data-gen-data.xlsx'),1))

  if(dir.exists('um')==FALSE){dir.create('um')}

  if(dir.exists('1k')==FALSE){dir.create('1k')}
  if(dir.exists('bum')==FALSE){dir.create('bum')}
  if(dir.exists('bma')==FALSE){dir.create('bma')}
  if(dir.exists('bma/occams window')==FALSE){dir.create('bma/occams window')}

  if(dir.exists('2k')==FALSE){dir.create('2k')}
  if(dir.exists('bmm2')==FALSE){dir.create('bmm2')}
  if(dir.exists('bmma2')==FALSE){dir.create('bmma2')}
  if(dir.exists('bmma2/occams window')==FALSE){dir.create('bmma2/occams window')}

  if(dir.exists('3k')==FALSE){dir.create('3k')}
  if(dir.exists('bmm3')==FALSE){dir.create('bmm3')}
  if(dir.exists('bmma3')==FALSE){dir.create('bmma3')}
  if(dir.exists('bmma3/occams window')==FALSE){dir.create('bmma3/occams window')}

  #modeling

```

## Lampiran 1. (Lanjutan)

```
run(aslidata,name=paste0(tmpid,'_asli')) #original data
for(j in 1:nsim){ #simulation
    run(gendata[j],name=paste0(tmpid,'_sim',j),plot=FALSE,append=TRUE)
}

cat('Proses rata-rata simulasi...\n')
###simulation averaging
##ks to excel
namrow <- c('asli',paste0('sim',1:nsim),'rata-rata')
ks <- as.matrix(read.table('ks.txt',header=TRUE))
ks <- rbind(ks,colMeans(ks[-1,]))
rownames(ks) <- namrow
write.xlsx(ks,'ks.xlsx')
file.remove('ks.txt')

##um to excel
cat('\ntrata-rata normal unimodal konvensional...\n')
um_mean <- as.matrix(read.table('um/um_mean.txt',header=TRUE))
ummod <- proc_bmodel(um_mean[-1,])
proc_plotmodel(y=aslidata,model=ummod,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
ggsave(filename=paste0('um/',tmpid,'_rata2_um.wmf'),width=14.07,height=12.85,units='cm')
um_mean <- rbind(um_mean,colMeans(um_mean[-1,]))
rownames(um_mean) <- namrow
write.xlsx(um_mean,'um/um_mean.xlsx')
file.remove('um/um_mean.txt')

##bum to excel
cat('\ntrata-rata bum...\n')
#component
bum_comp <- as.matrix(read.table('bum/bum_component.txt',header=TRUE))
bum_comp <- rbind(bum_comp,colMeans(bum_comp[-1,]))
rownames(bum_comp) <- namrow
write.xlsx(bum_comp,'bum/bum_component.xlsx')
file.remove('bum/bum_component.txt')

#hpd
bum_hpd <- as.matrix(read.table('bum/bum_hpd.txt',header=TRUE))
bum_hpd <- rbind(bum_hpd,colMeans(bum_hpd[-1,]))

rownames(bum_hpd) <- namrow
write.xlsx(bum_hpd,'bum/bum_hpd.xlsx')
file.remove('bum/bum_hpd.txt')

#mean_par
bum_mean <- as.matrix(read.table('bum/bum_mean.txt',header=TRUE))
bummod <- proc_bmodel(bum_mean[-1,])
proc_plotmodel(y=aslidata,model=bummod,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
ggsave(filename=paste0('bum/',tmpid,'_rata2_bum.wmf'),width=14.07,height=12.85,units='cm')
bum_mean <- rbind(bum_mean,colMeans(bum_mean[-1,]))
rownames(bum_mean) <- namrow
write.xlsx(bum_mean,'bum/bum_mean.xlsx')
file.remove('bum/bum_mean.txt')

##bma to excel
cat('\ntrata-rata bma...\n')
#component
bma_comp <- as.matrix(read.table('bma/bma_component.txt',header=TRUE))
bma_comp <- rbind(bma_comp,colMeans(bma_comp[-1,]))
rownames(bma_comp) <- namrow
write.xlsx(bma_comp,'bma/bma_component.xlsx')
file.remove('bma/bma_component.txt')

#hpd
bma_hpd <- as.matrix(read.table('bma/bma_hpd.txt',header=TRUE))
bma_hpd <- rbind(bma_hpd,colMeans(bma_hpd[-1,]))
rownames(bma_hpd) <- namrow
write.xlsx(bma_hpd,'bma/bma_hpd.xlsx')
```

## Lampiran 1. (Lanjutan)

```
file.remove('bma/bma_hpd.txt')

#mean_par
bma_mean <- as.matrix(read.table('bma/bma_mean.txt',header=TRUE))
bmamod <- proc_bmodel(bma_mean[-1,])
proc_plotmodel(y=aslidata,model=bmamod,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
ggsave(filename=paste0('bma/',tmpid,'_rata2_bma.wmf'),width=14.07,height=12.85,units='cm')
bma_mean <- rbind(bma_mean,colMeans(bma_mean[-1,]))
rownames(bma_mean) <- namrow
write.xlsx(bma_mean,'bma/bma_mean.xlsx')
file.remove('bma/bma_mean.txt')

##bmm2 to excel
cat("\trata-rata bmm2...\n")
#component
bmm2_comp <- as.matrix(read.table('bmm2/bmm2_component.txt',header=TRUE))
bmm2_comp <- rbind(bmm2_comp,colMeans(bmm2_comp[-1,]))
rownames(bmm2_comp) <- namrow
write.xlsx(bmm2_comp,'bmm2/bmm2_component.xlsx')
file.remove('bmm2/bmm2_component.txt')

#hpd
bmm2_hpd <- as.matrix(read.table('bmm2/bmm2_hpd.txt',header=TRUE))
bmm2_hpd <- rbind(bmm2_hpd,colMeans(bmm2_hpd[-1,]))
rownames(bmm2_hpd) <- namrow
write.xlsx(bmm2_hpd,'bmm2/bmm2_hpd.xlsx')
file.remove('bmm2/bmm2_hpd.txt')

#mean_par
bmm2_mean <- as.matrix(read.table('bmm2/bmm2_mean.txt',header=TRUE))
bmm2mod <- proc_bmodel(bmm2_mean[-1,])
proc_plotmodel(y=aslidata,model=bmm2mod,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
ggsave(filename=paste0('bmm2/',tmpid,'_rata2_bmm2.wmf'),width=14.07,height=12.85,units='cm')
bmm2_mean <- rbind(bmm2_mean,colMeans(bmm2_mean[-1,]))
rownames(bmm2_mean) <- namrow
write.xlsx(bmm2_mean,'bmm2/bmm2_mean.xlsx')
file.remove('bmm2/bmm2_mean.txt')

##bmma2 to excel
cat("\trata-rata bmma2...\n")
#component
bmma2_comp <- as.matrix(read.table('bmma2/bmma2_component.txt',header=TRUE))
bmma2_comp <- rbind(bmma2_comp,colMeans(bmma2_comp[-1,]))
rownames(bmma2_comp) <- namrow
write.xlsx(bmma2_comp,'bmma2/bmma2_component.xlsx')
file.remove('bmma2/bmma2_component.txt')

#hpd
bmma2_hpd <- as.matrix(read.table('bmma2/bmma2_hpd.txt',header=TRUE))
bmma2_hpd <- rbind(bmma2_hpd,colMeans(bmma2_hpd[-1,]))
rownames(bmma2_hpd) <- namrow
write.xlsx(bmma2_hpd,'bmma2/bmma2_hpd.xlsx')
file.remove('bmma2/bmma2_hpd.txt')

#mean_par
bmma2_mean <- as.matrix(read.table('bmma2/bmma2_mean.txt',header=TRUE))
bmma2mod <- proc_bmodel(bmma2_mean[-1,])
proc_plotmodel(y=aslidata,model=bmma2mod,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
ggsave(filename=paste0('bmma2/',tmpid,'_rata2_bmma2.wmf'),width=14.07,height=12.85,units='cm')
bmma2_mean <- rbind(bmma2_mean,colMeans(bmma2_mean[-1,]))
rownames(bmma2_mean) <- namrow
write.xlsx(bmma2_mean,'bmma2/bmma2_mean.xlsx')
file.remove('bmma2/bmma2_mean.txt')

##bmm3 to excel
cat("\trata-rata bmm3...\n")
#component
```

## Lampiran 1. (Lanjutan)

```
bmm3_comp <- as.matrix(read.table('bmm3/bmm3_component.txt',header=TRUE))
bmm3_comp <- rbind(bmm3_comp,colMeans(bmm3_comp[-1,]))
rownames(bmm3_comp) <- namrow
write.xlsx(bmm3_comp,'bmm3/bmm3_component.xlsx')
file.remove('bmm3/bmm3_component.txt')

#hpd
bmm3_hpd <- as.matrix(read.table('bmm3/bmm3_hpd.txt',header=TRUE))
bmm3_hpd <- rbind(bmm3_hpd,colMeans(bmm3_hpd[-1,]))
rownames(bmm3_hpd) <- namrow
write.xlsx(bmm3_hpd,'bmm3/bmm3_hpd.xlsx')
file.remove('bmm3/bmm3_hpd.txt')

#mean_par
bmm3_mean <- as.matrix(read.table('bmm3/bmm3_mean.txt',header=TRUE))
bmm3mod <- proc_bmodel(bmm3_mean[-1,])
proc_plotmodel(y=asldata,model=bmm3mod,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
ggsave(filename=paste0('bmm3/',tmpid,'_rata2_bmm3.wmf'),width=14.07,height=12.85,units='cm')
bmm3_mean <- rbind(bmm3_mean,colMeans(bmm3_mean[-1,]))
rownames(bmm3_mean) <- namrow
write.xlsx(bmm3_mean,'bmm3/bmm3_mean.xlsx')
file.remove('bmm3/bmm3_mean.txt')

##bmma3 to excel
cat("\trata-rata bmma3...\n")
#component
bmma3_comp <- as.matrix(read.table('bmma3/bmma3_component.txt',header=TRUE))
bmma3_comp <- rbind(bmma3_comp,colMeans(bmma3_comp[-1,]))
rownames(bmma3_comp) <- namrow
write.xlsx(bmma3_comp,'bmma3/bmma3_component.xlsx')
file.remove('bmma3/bmma3_component.txt')

#hpd
bmma3_hpd <- as.matrix(read.table('bmma3/bmma3_hpd.txt',header=TRUE))
bmma3_hpd <- rbind(bmma3_hpd,colMeans(bmma3_hpd[-1,]))
rownames(bmma3_hpd) <- namrow
write.xlsx(bmma3_hpd,'bmma3/bmma3_hpd.xlsx')
file.remove('bmma3/bmma3_hpd.txt')

#mean_par
bmma3_mean <- as.matrix(read.table('bmma3/bmma3_mean.txt',header=TRUE))
bmma3mod <- proc_bmodel(bmma3_mean[-1,])
proc_plotmodel(y=asldata,model=bmma3mod,main="",xlab='Perbedaan Ekspresi Gen',ylab='Densitas')
ggsave(filename=paste0('bmma3/',tmpid,'_rata2_bmma3.wmf'),width=14.07,height=12.85,units='cm')
bmma3_mean <- rbind(bmma3_mean,colMeans(bmma3_mean[-1,]))
rownames(bmma3_mean) <- namrow
write.xlsx(bmma3_mean,'bmma3/bmma3_mean.xlsx')
file.remove('bmma3/bmma3_mean.txt')
cat(paste('Simulasi',tmpid,'Selesai.\n'))

akhir <- Sys.time()
waktu <- akhir-awal
sink('run_time.txt')
print(awal)
print(akhir)
print(waktu)
sink()
}
```

### **BUGS TOOLS.R:**

```
proc_takebugs <- function(bug){
#-----
#Construct bugs sample estimation to matrix
#
#input
#bug: bug object from bugs function (R2OpenBUGS)
#output
```



## Lampiran 1. (Lanjutan)

```
#sample parameters from bayes estimation
#-----
      params <- bug
      pD <- params[['pD']]
      deviance <- params[['sims.list']][['deviance']]
      lambda <- params[['sims.list']][['lambda']]
      tau <- params[['sims.list']][['tau']]
      sigma <- sqrt(1/tau)
      if(is.null(ncol(lambda))) {
        w <- rep(1,length(lambda))
        ncomp <- 1
      } else {
        w <- params[['sims.list']][['w']]
        ncomp <- ncol(w)
      }
      result <- cbind(pD,deviance,w,lambda,tau,sigma)
      colnames(result) <- c('pD','deviance',paste0('w.',1:ncomp),paste0('lambda.',1:ncomp),
        paste0('tau.',1:ncomp),paste0('sigma.',1:ncomp))
      return(result)
}
```

### MODEL TOOLS.R:

```
proc_model0 <- function(y){
#-----
#Create normal conventional object
#
#input
#y: data sample vector
#output
#      conventional normal pdf object
#-----
      ncomp <- 1
      w <- 1
      lambda <- mean(y)
      sigma <- sd(y)
      param <- list(ncomp=ncomp,w=w,lambda=lambda,sigma=sigma)

      normpdf <- proc_normpdf(param)
      out <- list(type=0,parameters=param,pdf=normpdf)
      return(out)
}

par_bmodel <- function(takebugs){
#-----
#Construct bayesian normal parameter to list object
#
#input
#      takebugs          :          object from proc_takebugs
#output
#      parameters of normal distribution
#-----
      name <- colnames(takebugs)
      param <- colMeans(takebugs)
      idxw <- grep('w.',name)
      ncomp <- length(idxw)
      idxlambda <- grep('lambda.',name)
      idxsigma <- grep('sigma.',name)
      return(list(ncomp=ncomp,w=param[idxw],lambda=param[idxlambda],
        sigma=param[idxsigma]))
}

proc_bmodel <- function(takebugs){
#-----
#Construct bayesian normal object
#
#input
```

## Lampiran 1. (Lanjutan)

```
#      takebugs      :      object from proc_takebugs
#output
#      bayesian normal pdf object
#-----
      param <- par_bmodel(takebugs)
      normpdf <- proc_normpdf(param)
      out <- list(type=1,parameters=param,pdf=normpdf)
      return(out)
}

par_bAvmodel <- function(occams){
#-----
#Construct bayesian normal averaging parameter to list object
#
#input
#      occams      :      occams object from proc_occamswindow
#output
#      parameters of normal averaging model
#-----
      name <- colnames(occams)
      idxw <- grep('w.',name)
      ncomp <- length(idxw)
      idxlambda <- grep('lambda.',name)
      idxsigma <- grep('sigma.',name)
      p <- occams[, 'PMkIx']
      nmodel <- length(p)
      w <- matrix(occams[,idxw],ncol=ncomp)
      lambda <- matrix(occams[,idxlambda],ncol=ncomp)
      sigma <- matrix(occams[,idxsigma],ncol=ncomp)
      return(list(nmodel=nmodel,ncomp=ncomp,p=p,w=w,
                  lambda=lambda,sigma=sigma))
}

proc_bAvmodel <- function(occams){
#-----
#Construct bayesian normal averaging object
#
#input
#      estbugs      :      object from proc_estbugs
#output

#      bayesian normal averaging pdf object
#-----
      param <- par_bAvmodel(occams)
      normpdf <- proc_normAvpdf(param)
      out <- list(type=2,parameters=param,pdf=normpdf)
      return(out)
}

proc_occamswindow <- function(takebugs){
#-----
#Occams window selection procedure
#
#input
#      takebugs      :      object from proc_takebugs
#output
#      sample selected by occam's window
#-----
      name <- colnames(takebugs)
      colw <- grep('w.',name)
      colambda <- grep('lambda.',name)
      coltau <- grep('tau.',name)
      colsigma <- grep('sigma.',name)
      idxcol <- c(colw,colambda,coltau,colsigma)
      dev <- takebugs[, 'deviance']
      pD <- takebugs[, 'pD']
      BIC <- dev+(pD*log10(3))
      i <- 1:nrow(takebugs)
      repeat{
```

## Lampiran 1. (Lanjutan)

```

        BIC <- BIC[i]
        maxBIC <- max(BIC)
        Numerator <- exp(-.5*(BIC-maxBIC))
        Denominator <- sum(Numerator)
        PMkIx <- Numerator/Denominator
        maxPMkIx <- max(PMkIx)
        A <- maxPMkIx/PMkIx
        ow.test <- as.numeric(A<=20)
        if(all(ow.test==1)){
            break
        }
        i <- which(ow.test==1)
    }
    accept <- cbind(i,dev[i],pD,BIC,takebugs[i,idxcol],maxBIC,Numerator,
                    Denominator,PMkIx,maxPMkIx,A,ow.test)
    namecol <- c('iteration','deviance','pD','BIC',name[idxcol],'maxBIC',
                'Numerator','Denominator','PMkIx','maxPMkIx','ratio','test')
    colnames(accept) <- namecol
    rownames(accept) <- paste0(1:nrow(accept))
    return(accept)
}

proc_plotmodel <- function(y,model,main=NULL,xlab='Y',ylab='Density',
                           xlimit=NULL,color='red'){
#-----
#Plot normal pdf model
#
#require(ggplot2)
#input
#      y      :      vector data sample
#      model   :      pdf take from proc_model0, proc_bmodel,
#                      proc_bAvmodel
#output
#      pdf curve plot
#-----
    type <- model[['type']]
    ncomp <- model[['parameters']][['ncomp']]
    pdfn <- model[['pdfn']]
    if(is.null(main)){
        plotname <- c('Conventional','BayesNonAveraging','BayesAveraging')
        main <- paste(plotname[type+1],'Normal',ncomp,'Component')

    }else{
        main <- main
    }

    if(is.null(xlimit)){
        y <- as.numeric(y)
        mu <- mean(y)
        sig <- sd(y)
        lower <- floor(mu-5*sig)
        upper <- ceiling(mu+5*sig)
    }else{
        lower <- floor(xlimit[1])
        upper <- ceiling(xlimit[2])
    }
    ggplot()+stat_function(aes(x=lower:upper),size=0.65,colour=color,fun=pdfn,n=200)+
        ggtitle(main)+xlab(xlab)+ylab(ylab)
}

proc_hpd <- function(y,model){
#-----
#Search hpd interval of normal model
#
#input
#      y      :      data sample vector
#      model   :      pdf take from proc_model0, proc_bmodel,
#                      proc_bAvmodel

```

## Lampiran 1. (Lanjutan)

```
#output
#      results of hpd
#-----
pdfn <- model[['pdfn']]
a <- min(y)
b <- max(y)
syseq <- function(x){
  fx <- numeric(2)
  fx[1] <- integrate(pdfn,x[1],x[2])$value-.95
  fx[2] <- pdfn(x[2])-pdfn(x[1])
  return(fx)
}
x0 <- sort(runif(2,a,b))
term <- 6
k <- 0
while(term!=1){
  k <- k+1
  est <- nleqslv(x0,syseq.method='Newton')
  term <- est$termcd
  x0 <- sort(runif(2,a,b))
}
output <- c(est[['x']],x0,est[['fvec']],est[['iter']],k)
names(output) <- c('lower','upper','lower0','upper0',
                  'f1val','f2val','iter','ntrial')

return(output)
}

proc_component <- function(model){
#-----
#Calculate pdf component (mean, var)
#
#input
#      model      :      pdf take from proc_model0, proc_bmodel,
#                      proc_bAvmodel
#output
#      mean and var component
#-----
type <- model[['type']]
param <- model[['parameters']]
ncomp <- param[['ncomp']]
w <- param[['w']]
lambda <- param[['lambda']]
sigma <- param[['sigma']]

if(type==0|type==1){
  lambda2 <- lambda^2
  sigma2 <- sigma^2
  meanf <- sum(w*lambda)
  varf <- sum(w*sigma2)+sum(w*lambda2)-(sum(w*lambda))^2
}else if(type==2){
  p <- param[['p']]
  mean_ymi <- rowSums(w*lambda)
  mean_comp <- p*mean_ymi
  meanf <- sum(mean_comp)
  mean_ymicor2 <- (mean_ymi-meanf)^2
  lambdacor2 <- (lambda-meanf)^2
  sigma2 <- sigma^2
  lambdasigma <- lambdacor2+sigma2
  var_ymi <- rowSums(w*lambdasigma)
  var_comp <- p*mean_ymicor2+var_ymi
  varf <- sum(var_comp)
}
component <- c(meanf,varf)
names(component) <- c('meanf','varf')
return(component)
}
```

## Lampiran 1. (Lanjutan)

### PDF TOOLS.R:

```
proc_normpdf <- function(param){
#-----
#Create normal pdf R object function
#
#input
#      param          :      object from par_bmodel, list(w,lambda,sigma)
#output
#      normal pdf
#-----
  w <- param[['w']]
  lambda <- param[['lambda']]
  sigma <- param[['sigma']]
  ncomp <- length(w)
  normpdf <- function(x){
    f <- 0
    for (j in 1:ncomp){
      f <- f+w[j]*(1/sqrt(2*pi*sigma[j]^2))*exp(-.5*((x-lambda[j])/sigma[j])^2)
    }
    return(as.numeric(f))
  }
}

proc_normAvpdf <- function(owparam){
#-----
#Create normal averaging pdf R object function
#
#input
#      owparam          :      object from par_bAvmodel
#output
#      normal averaging pdf
#-----
  nmodel <- owparam[['nmodel']]
  ncomp <- owparam[['ncomp']]
  p <- owparam[['p']]
  w <- owparam[['w']]
  lambda <- owparam[['lambda']]
  sigma <- owparam[['sigma']]
  normAvpdf <- function(x){
    fx <- 0
    for(i in 1:nmodel){
      f <- 0
      for (j in 1:ncomp){
        f <- f+w[i,j]*(1/sqrt(2*pi*sigma[i,j]^2))*exp(-.5*((x-
lambda[i,j])/sigma[i,j])^2)
      }
      fx <- fx+p[i]*f
    }
    return(as.numeric(fx))
  }
}

proc_rmixnorm <- function(n,param){
#-----
#Draw random sample of mixture normal
#
#input
#      n                :      number of random to draw
#      param            :      parameters of pdf (list(w=...,lambda=...,sigma=...))
#output
#      random number
#-----
  w <- param[['w']]
  lambda <- param[['lambda']]
  sigma <- param[['sigma']]
  if(n==1){
    k <- sample(length(lambda),size=1,prob=w)
```

## Lampiran 1. (Lanjutan)

```

    }else{
      y <- rnorm(1,mean=lambda[k],sd=sigma[k])
    }
    rmix <- function(k){rnorm(nk[k],mean=lambda[k],sd=sigma[k])}
    nk <- as.vector(rmultinom(n=1,size=n,prob=w))
    y <- sample(unlist(lapply(seq_along(nk),rmix)))
  }
  return(y)
}

proc_ksnorm <- function(y,model,alpha=0.05){
#-----
#Kolmogorov Smirnov test for normal model
#
#input
#      y      :      data sample vector
#      model   :      pdf take from proc_model0, proc_bmodel,
#                   proc_bAvmodel
#      alpha   :      signif level(0.10, 0.05, 0.01)
#output
#      ks test result (ks, val(0:H0,1:H1 -> accept))
#-----
  pdfn <- model[['pdfn']]
  y <- sort(y)
  n <- length(y)
  epdf <- prop.table(table(y))
  S <- F0 <- Df <- rep(0,n)
  Df1 <- 0
  for(i in 1:n){
    S[i] <- sum(epdf[1:i])
    F0[i] <- integrate(pdfn,-Inf,y[i])$value
    Df[i] <- abs(S[i]-F0[i])
    if(i==1){
      Df1[i] <- F0[1]
    }else{
      Df1[i] <- abs(S[i-1]-F0[i])
    }
  }
  D <- max(Df,Df1)
  #return(D)

  namequant <- c('p0.90'=0.10,'p0.95'=0.05,'p0.99'=0.01)
  denumquant <- c(1.22,1.36,1.63)
  idx <- which(namequant==alpha)
  n <- length(y)
  if(n<=40){
    Dtab <- Quantileks[n,idx]
  }else{
    Dtab <- denumquant[idx]/sqrt(n)
  }
  test <- c(D,as.numeric(D>Dtab))
  names(test) <- c('Statistis','Decision')
  return(test)
}

# proc_kstest <- function(y,model,alpha=0.05){
# #-----
# #input
# #      y      :      data sample vector
# #      model   :      pdf take from proc_model0, proc_bmodel,
# #                   proc_bAvmodel
# #      alpha   :      signif level(0.10, 0.05, 0.01)
# #output
# #      ks test result (ks, val(0:H0,1:H1 -> accept))
# #-----
#   namequant <- c('p0.90'=0.10,'p0.95'=0.05,'p0.99'=0.01)
#   denumquant <- c(1.22,1.36,1.63)
#   idx <- which(namequant==alpha)
#   n <- length(y)

```

## Lampiran 1. (Lanjutan)

```
# D <- proc_ksnorm(y,model)
# if(n<=40){
#   # Dtab <- BayesNormAv:::Quantileks[n,idx]
# }else{
#   # Dtab <- denumquant[idx]/sqrt(n)
# }
# test <- c(D,as.numeric(D>Dtab))
# names(test) <- c('Statistcs','Decision')
# return(test)
# }
```

**Lampiran 2.** Fungsi ID Gen dalam Jaringan Tanaman *Chickpea* (Harijati, 2007)

No	ID Gen	Fungsi Umum	Fungsi Spesifik Dugaan
1	LS0024	<i>Defence</i>	<i>Dis.resist response prot. DRRG49-C</i>
2	LS0035	<i>Defence</i>	<i>Putative auxin-repressed prot.</i>
3	LS0616	<i>Defence</i>	<i>Putative auxin-repressed prot.</i>
4	LS0159	<i>Defence</i>	<i>Beta-glucan binding protein</i>
5	LS0162	<i>Defence</i>	<i>Pathogenesis related prot.</i>
6	LS0185	<i>Defence</i>	<i>Pathogenesis related prot.</i>
7	LS0612	<i>Defence</i>	<i>Laccase-like protein</i>
8	LS0081	<i>Defence</i>	<i>Pathogenesis related prot. 4A</i>
9	LS0752	<i>Defence</i>	<i>Pathogenesis related prot. 4A</i>
10	LS0759	<i>Defence</i>	<i>Pathogenesis related prot. 4A</i>
11	LS0688	<i>Defence</i>	<i>Pathogenesis related prot. 4A</i>
12	U070	<i>Defence</i>	<i>Homol. To Avr9/Cf9 rapidly elic.prot.</i>
13	U273	<i>Defence</i>	<i>Prot.cont. leucine zipper motif</i>
14	U278	<i>Defence</i>	<i>Homol.to SNAKIN2 antimicrobial peptide</i>
15	U017	<i>Defence</i>	<i>PR protein class 10</i>
16	LS0132	<i>Energy</i>	<i>Thioredoxin</i>
17	LS0341	<i>Energy</i>	<i>Thioredoxin H-type 1</i>
18	U001	<i>Energy</i>	<i>Photosystem I reaction center subunit IX</i>
19	U325	<i>Energy</i>	<i>Photosystem I assembly prot.ycf3</i>
20	U385	<i>Energy</i>	<i>Chloroplast DNA</i>
21	U447	<i>Energy</i>	<i>ATP synthase C chain</i>
22	CA0426	<i>Energy</i>	<i>Chlorophyll a/b binding protein</i>
23	CA0534	<i>Energy</i>	<i>Chlorophyll a/b binding protein</i>
24	CA0702	<i>Energy</i>	<i>NADH-plastoquinone oxidoreductase chain 1</i>
25	CA0810	<i>Energy</i>	<i>Chloroplast val-tRNA</i>
26	CA1120	<i>Energy</i>	<i>Chloroplast DNA</i>
27	LS0403	<i>Cellular Communic/Signal Transduction</i>	<i>RAC-GTP binding protein-like</i>
28	LS0551	<i>Cellular Communic/Signal Transduction</i>	<i>Nonspecific lipid-transfer protein precursor</i>
29	LS0744	<i>Cellular Communic/Signal Transduction</i>	<i>Protein kinase-like protein</i>
30	U033	<i>Cellular Communic/Signal Transduction</i>	<i>Protein kinase</i>
31	LS0156	<i>Cell Rescue/Death/Aging</i>	<i>M.sativa environmental stress inducible protein</i>
32	LS0287	<i>Cell Rescue/Death/Aging</i>	<i>Metallothionein-like protein 1</i>
33	LS0691	<i>Cell Rescue/Death/Aging</i>	<i>Metallothionein-like protein 1</i>
34	U133	<i>Cell Rescue/Death/Aging</i>	<i>Phosphate-induced protein</i>
35	U187	<i>Cell Rescue/Death/Aging</i>	<i>Proline oxidase enzyme 1</i>
36	U212	<i>Cell Rescue/Death/Aging</i>	<i>Glutathione s-transferase enzyme</i>
37	U380	<i>Cell Rescue/Death/Aging</i>	<i>Supox.dismut.Cu chaper.prec.</i>



## Lampiran 2. (Lanjutan)

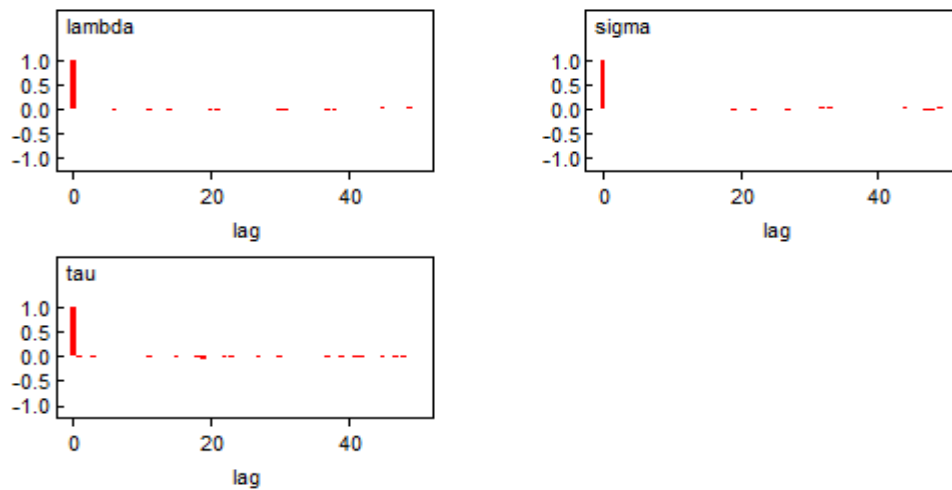
No	ID Gen	Fungsi Umum	Fungsi Spesifik Dugaan
38	LS0493	<i>Cellular Metabolism</i>	<i>Glycolate Oxidase</i>
39	LS0668	<i>Cellular Metabolism</i>	<i>Putat.Ubiquitin Protein</i>
40	LS0675	<i>Cellular Metabolism</i>	<i>Ubiquitin-conjug. Enzyme E2</i>
41	LS0801	<i>Cellular Metabolism</i>	<i>Cysteine Proteinase 15A Precursor</i>
42	LS0951	<i>Cellular Metabolism</i>	<i>Catalase</i>
43	U072	<i>Cellular Metabolism</i>	<i>Cytochrome C Oxidase subunit</i>
44	U110	<i>Cellular Metabolism</i>	<i>UDP-glucose 4-epimerase (EC 5.1.3.2)</i>
45	U502	<i>Cellular Metabolism</i>	<i>Cytochrome C Oxidase subunit</i>
46	U379	<i>Cellular Metabolism</i>	<i>Similar to endopeptidase</i>
47	CA0988	<i>Protein Synthesis</i>	<i>26S rRNA</i>
48	U111	<i>Protein Synthesis</i>	<i>18S nucl rRNA</i>
49	U337	<i>Cell Cycle and DNA Processing</i>	<i>RNA/ssDNA binding protein</i>
50	U407	<i>Transport Facilitation</i>	<i>Beta Adaptin like protein</i>
51	U471	<i>Transport Facilitation</i>	<i>DNAJ like protein</i>
52	U010	<i>Unknown or Unclear Function</i>	<i>Unknown</i>
53	U152	<i>Unknown or Unclear Function</i>	<i>Unknown</i>
54	U319	<i>Unknown or Unclear Function</i>	<i>Unknown</i>
55	U464	<i>Unknown or Unclear Function</i>	<i>Unknown</i>
56	U488	<i>Unknown or Unclear Function</i>	<i>Unknown</i>
57	CA0890	<i>Unknown or Unclear Function</i>	<i>Unknown</i>
58	U055	<i>Unknown or Unclear Function</i>	<i>Unclear</i>
59	U347	<i>Unknown or Unclear Function</i>	<i>Unclear</i>
60	U369	<i>Unknown or Unclear Function</i>	<i>Unclear</i>
61	U392	<i>Unknown or Unclear Function</i>	<i>Unclear</i>
62	CA0043	<i>Unknown or Unclear Function</i>	<i>Unclear</i>

**Lampiran 3.** Hasil *Running* Program BUGS untuk 7 ID Gen pada Model BUM-Normal Data Simulasi

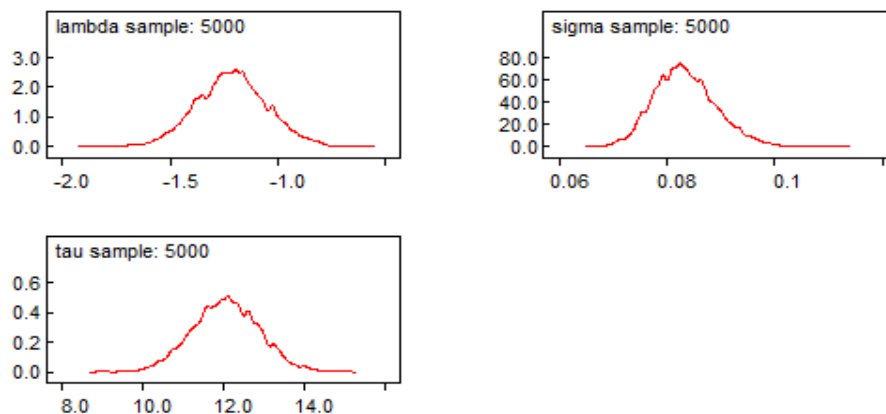
1. Rata-Rata Data Simulasi untuk ID Gen LS0612 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.2.1.** Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen LS0612

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-1,212	0,1643	0,001 9	-1,53	-1,212	-0,8811	1	5.000
sigma	0,0834	0,0058	7,8E- 05	0,0732	0,0830	0,0958	1	5.000
tau	12,05	0,8243	0,011 2	10,44	12,05	13,67	1	5.000

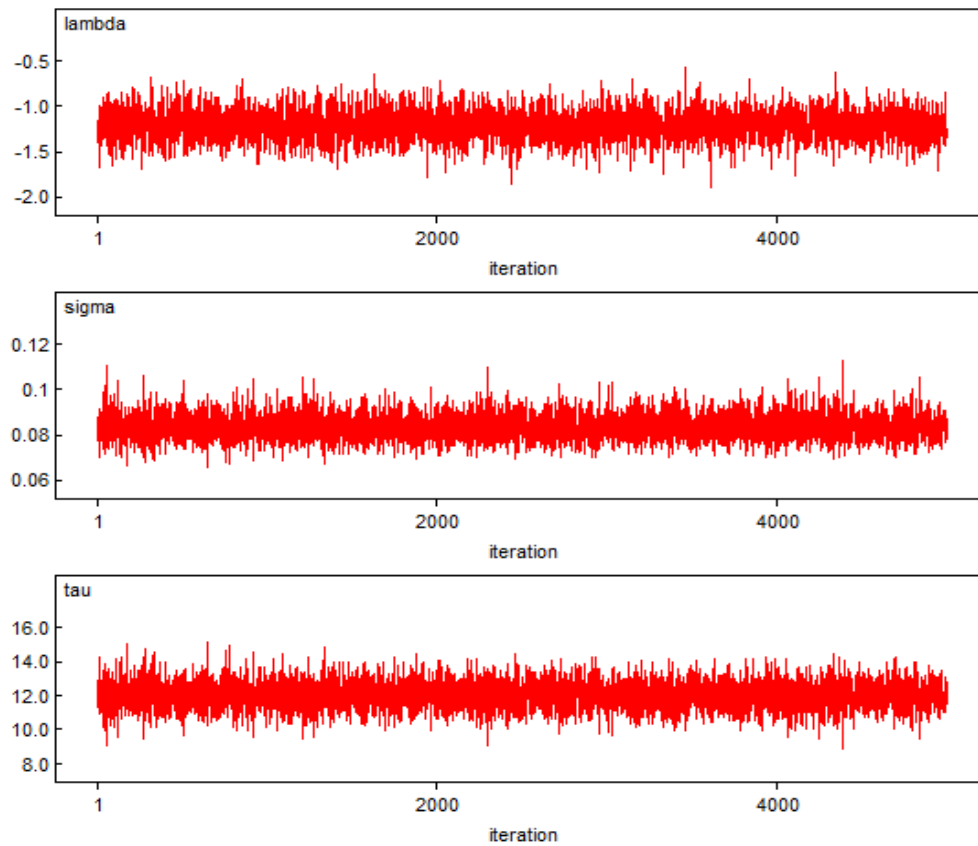


**Gambar L.2.1.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0612

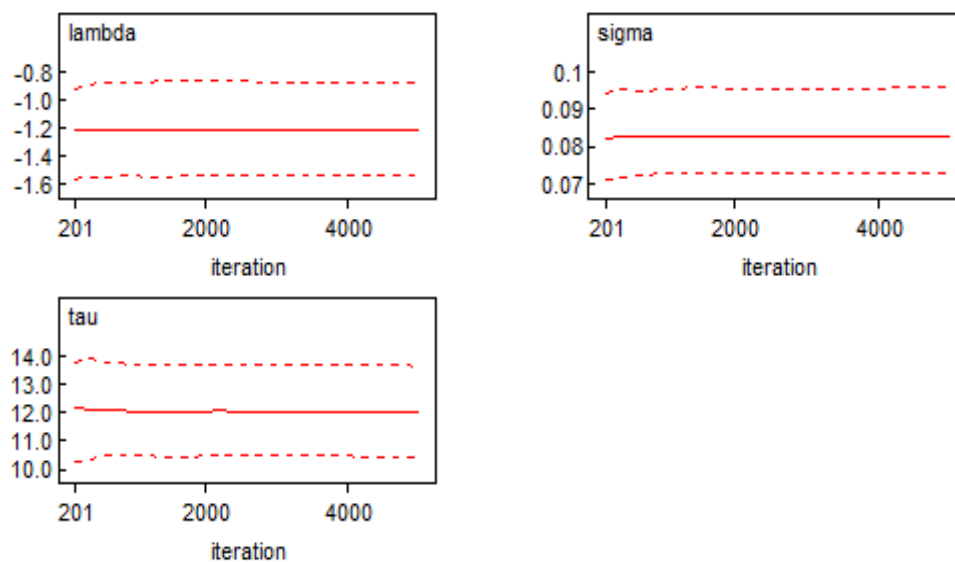


**Gambar L.2.2.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0612

### Lampiran 3. (Lanjutan)



**Gambar L.2.3.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0612



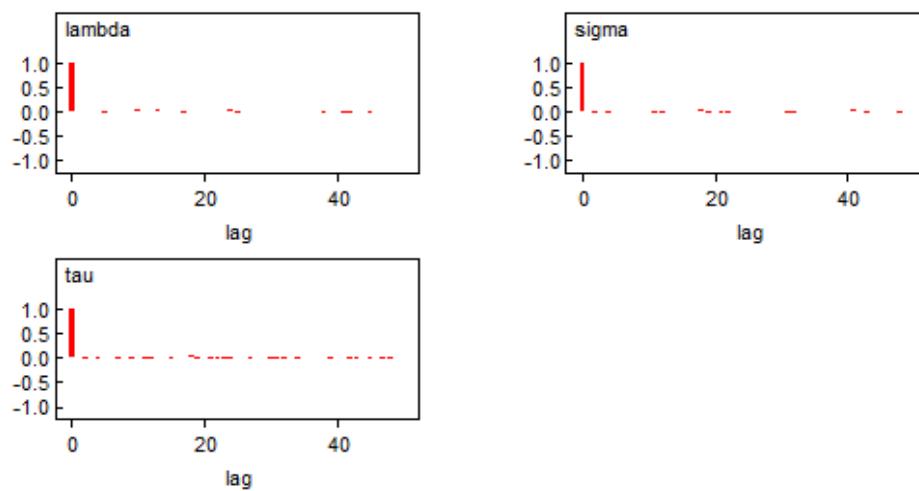
**Gambar L.2.4.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0612

### Lampiran 3. (Lanjutan)

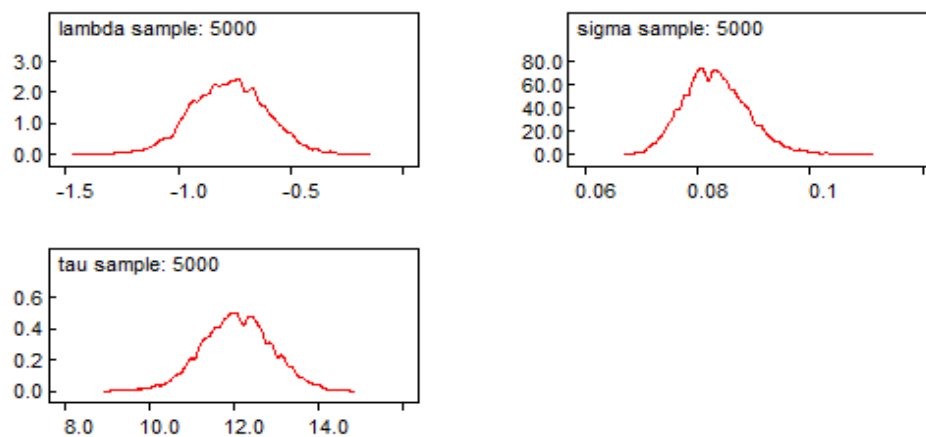
#### 2. Rata-Rata Data Simulasi untuk ID Gen LS0759 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.2.2.** Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen LS0759

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,773	0,1678	0,0026	-1,1	-0,7716	-0,4453	1	5.000
sigma	0,0832	0,0057	7,3E-02	0,0732	0,0829	0,0956	1	5.000
tau	12,07	0,8143	0,0101	10,47	12,06	13,66	1	5.000

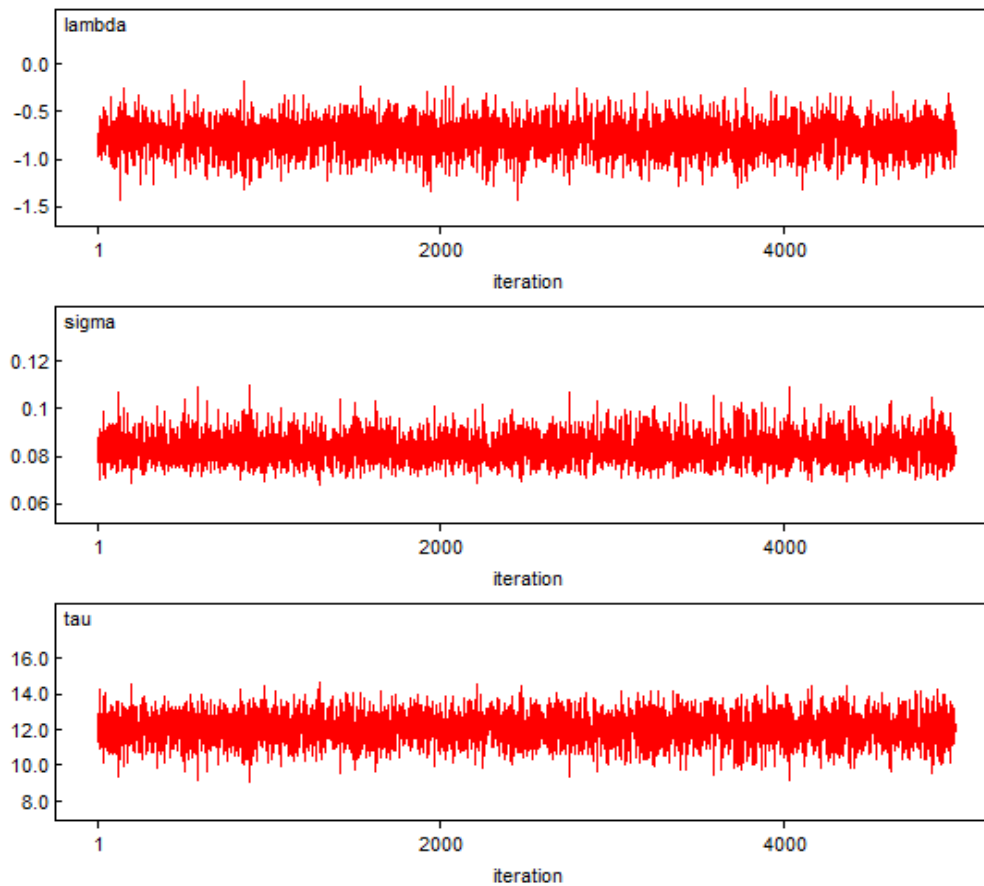


**Gambar L.2.5.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0759

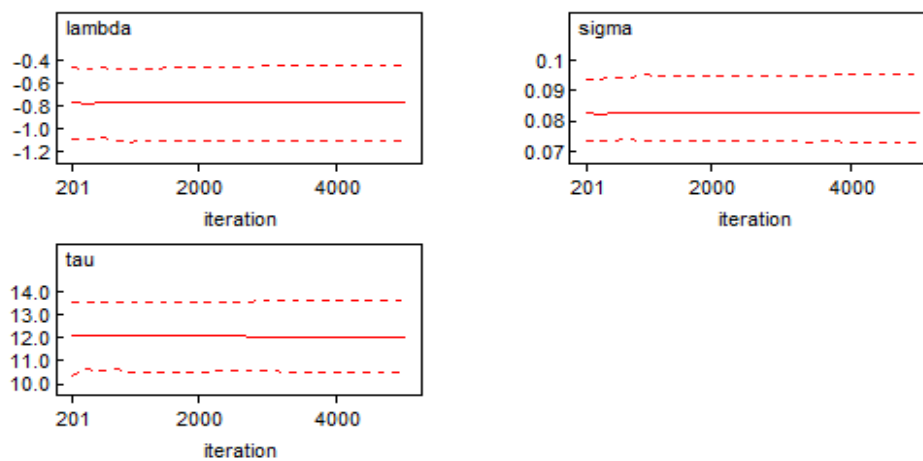


**Gambar L.2.6.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0759

### Lampiran 3. (Lanjutan)



**Gambar L.2.7.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0759



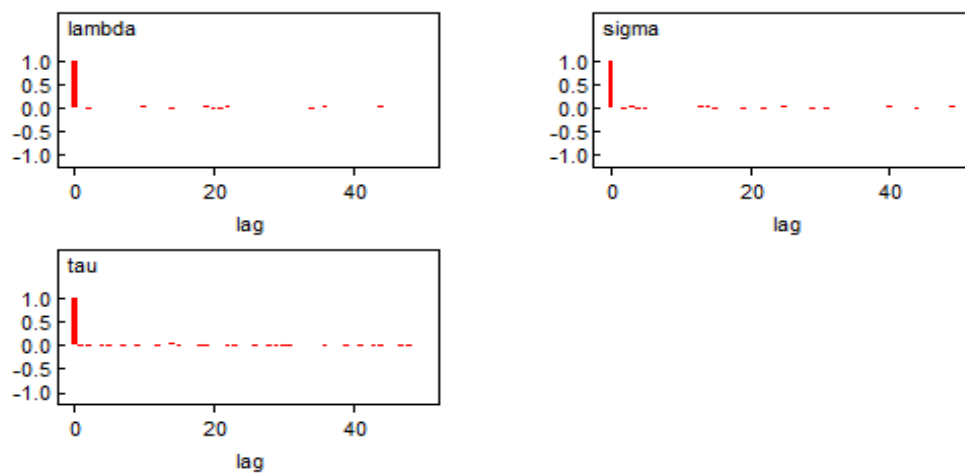
**Gambar L.2.7.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0759

### Lampiran 3. (Lanjutan)

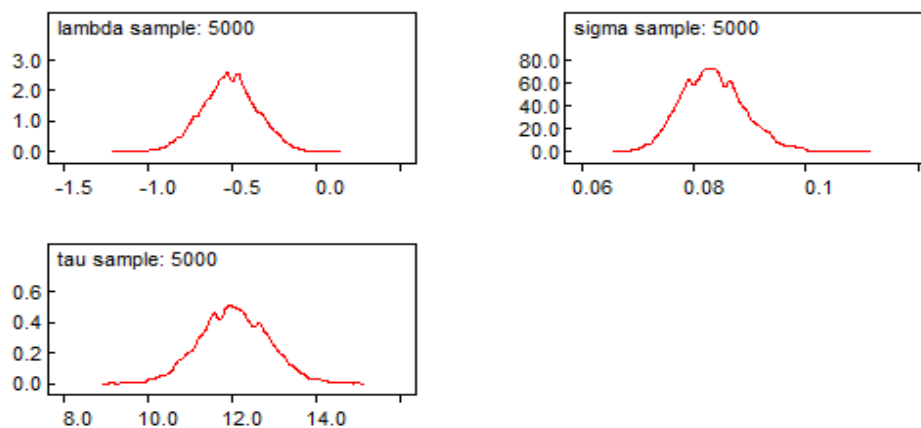
#### 3. Rata-Rata Data Simulasi untuk ID Gen U278 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.2.3.** Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen U278

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,5177	0,1662	0,0025	-0,8436	-0,5178	-0,1925	1	5.000
sigma	0,0835	0,0057	7,7E-02	0,0733	0,0832	0,0956	1	5.000
tau	12,04	0,8114	0,0109	10,47	12,03	13,64	1	5.000

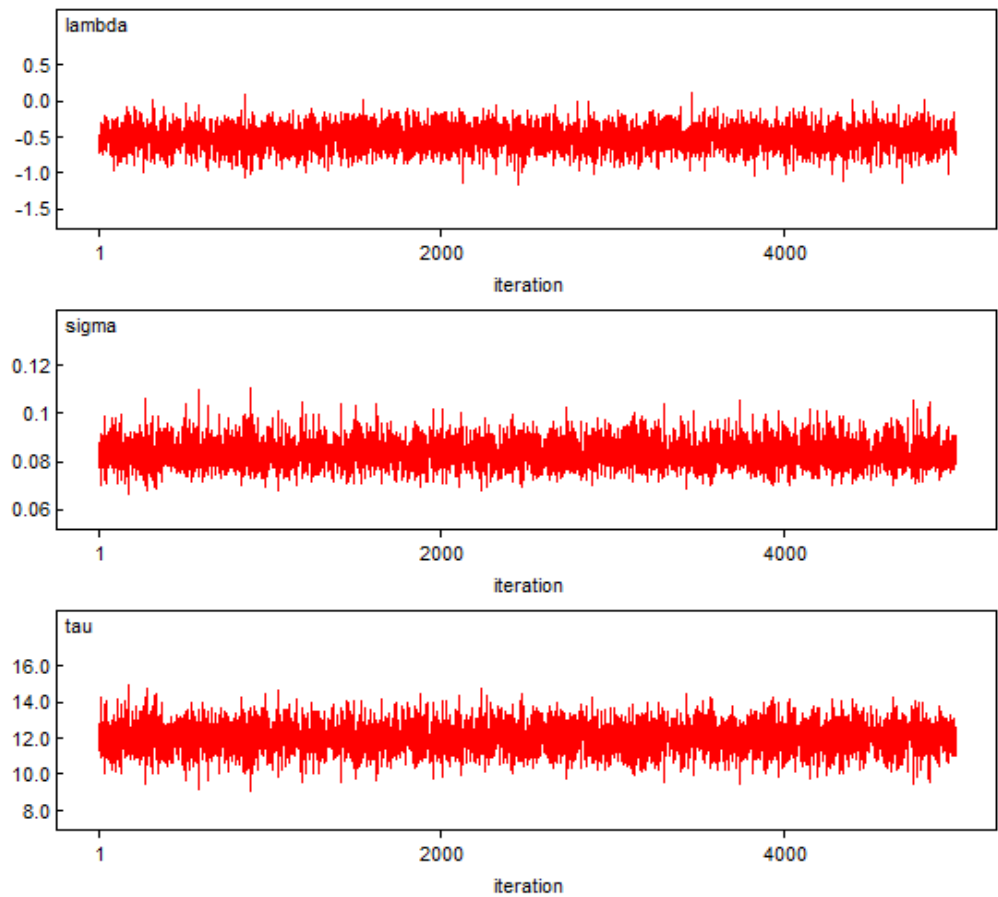


**Gambar L.2.8.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen U278

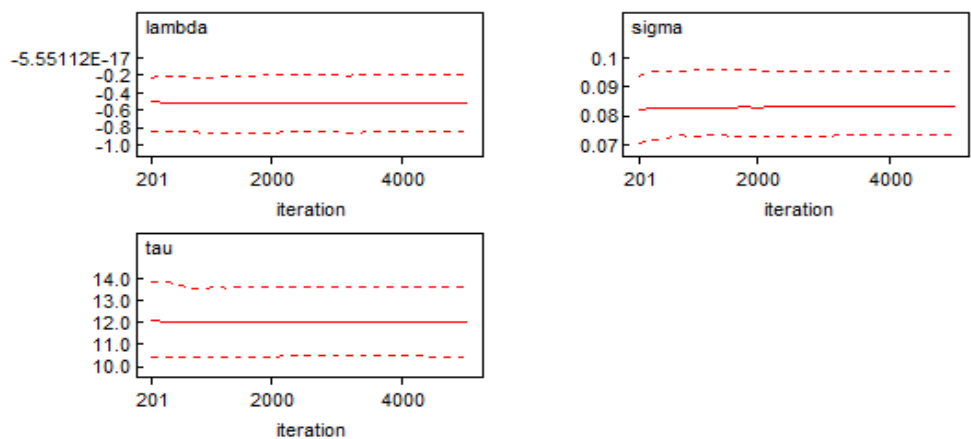


**Gambar L.2.9.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen U278

**Lampiran 3. (Lanjutan)**



**Gambar L.2.10.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen U278



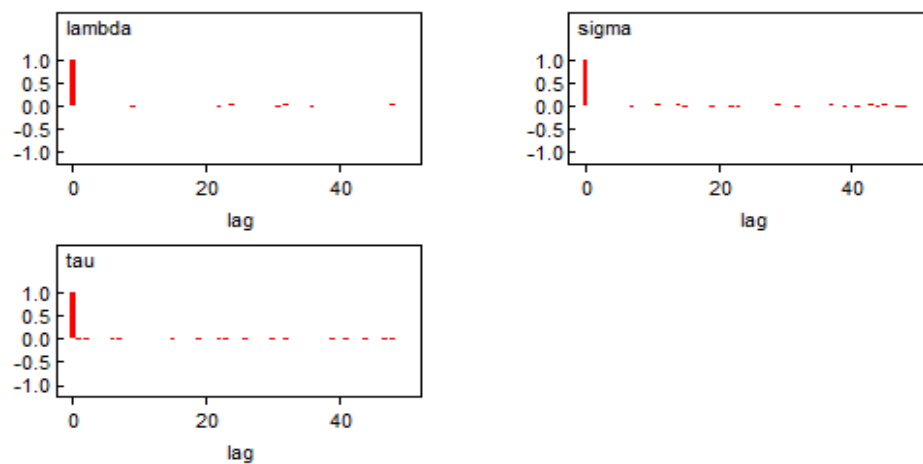
**Gambar L.2.11.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen U278

### Lampiran 3. (Lanjutan)

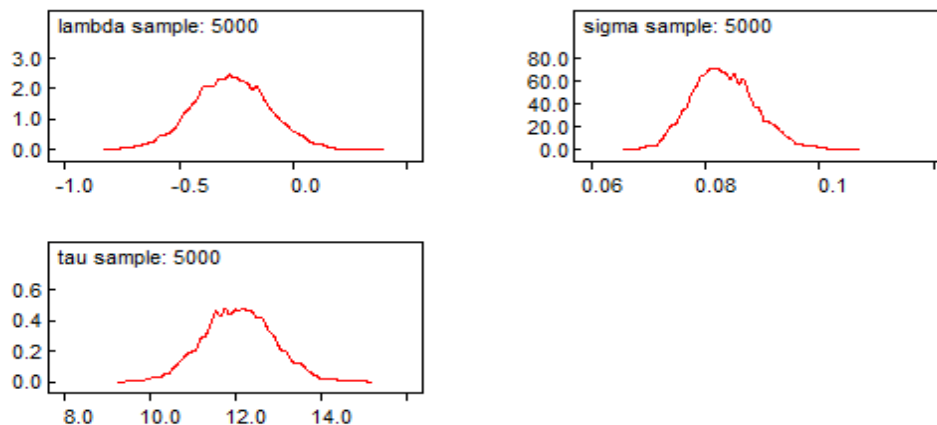
#### 4. Rata-Rata Data Simulasi untuk ID Gen U017 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.2.4.** Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen U017

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,2709	0,1654	0,0024	-0,5932	-0,2724	0,0549	1	5.000
sigma	0,0833	0,0057	8,8E-02	0,0732	0,0829	0,0951	1	5.000
tau	12,06	0,8104	0,0125	10,51	12,06	13,66	1	5.000



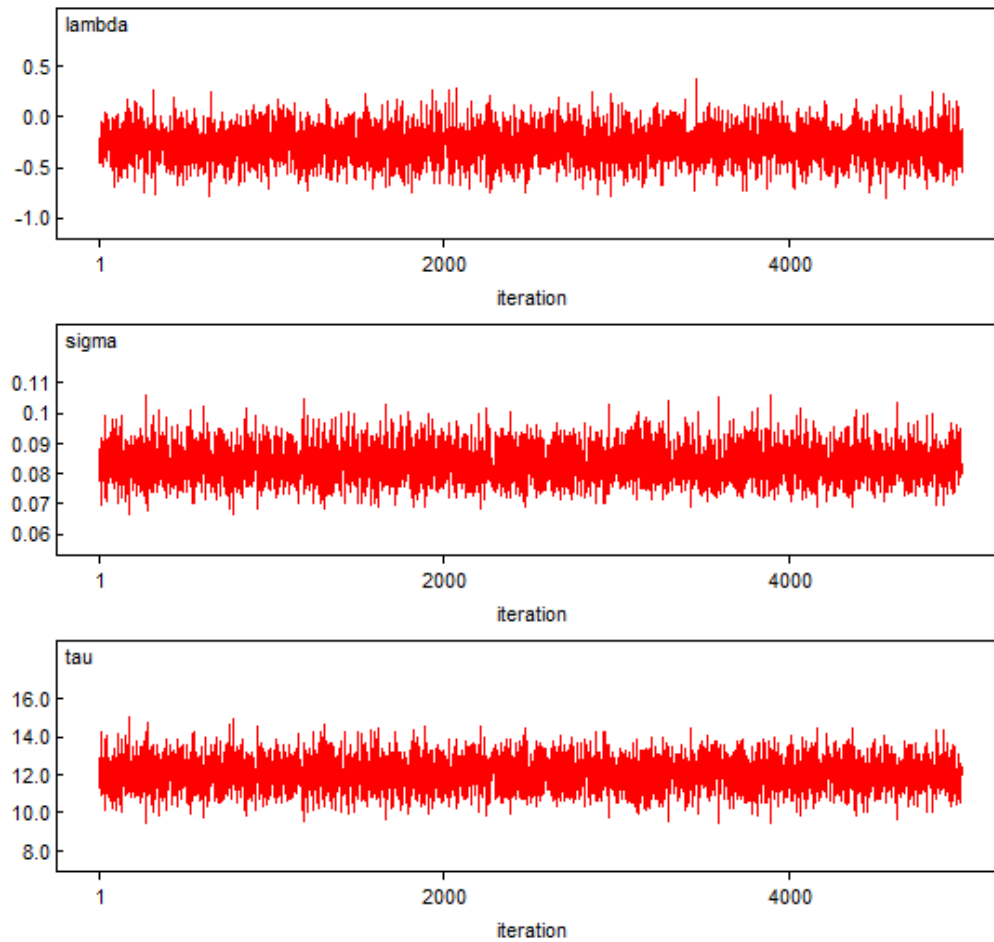
**Gambar L.2.12.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen U017



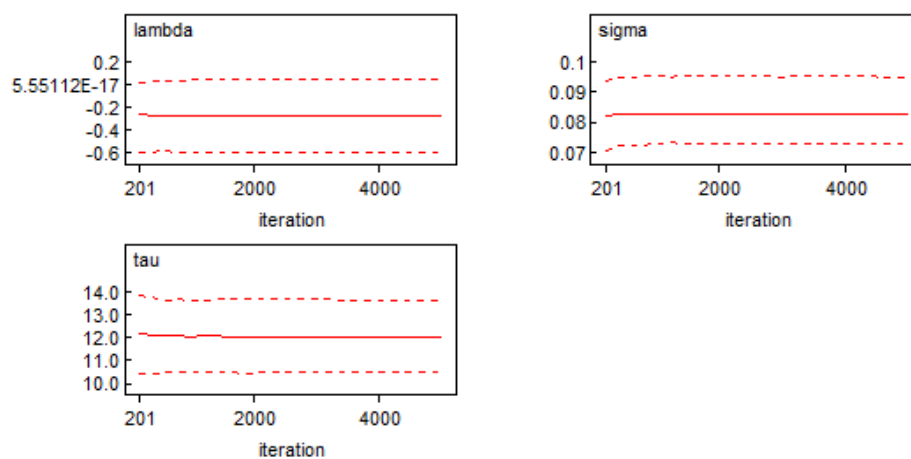
**Gambar L.2.13.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen U017



### Lampiran 3. (Lanjutan)



**Gambar L.2.14.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen U017



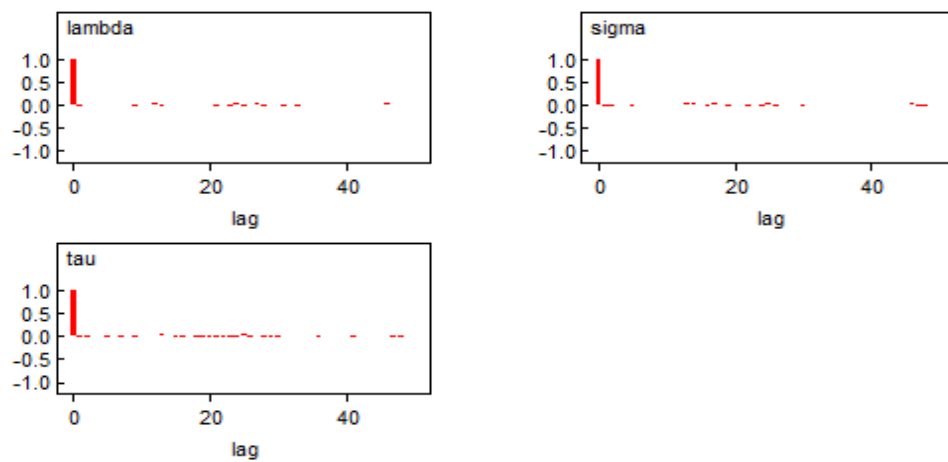
**Gambar L.2.15.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen U017

### Lampiran 3. (Lanjutan)

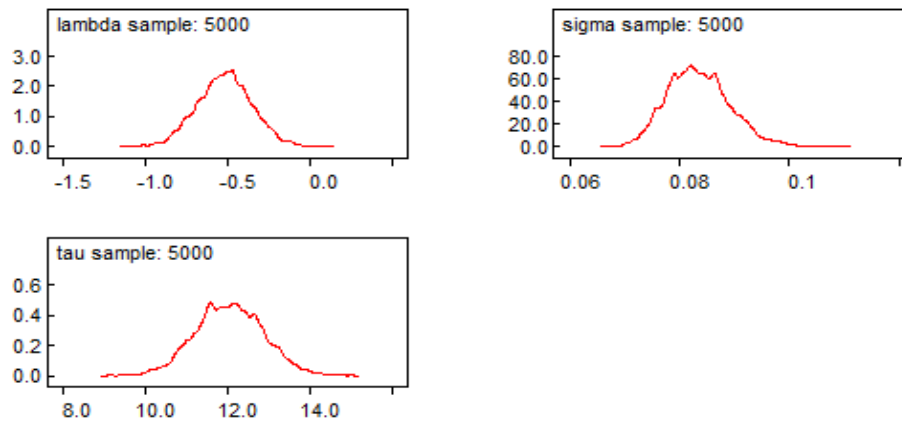
#### 5. Rata-Rata Data Simulasi untuk ID Gen LS0132 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.2.5.** Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen LS0132

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,5196	0,1669	0,0024	-0,8427	-0,5169	-0,1925	1	5.000
sigma	0,0835	0,0057	7,4E-05	0,0734	0,0831	0,0957	1	5.000
tau	12,03	0,8138	0,0104	10,45	12,03	13,63	1	5.000

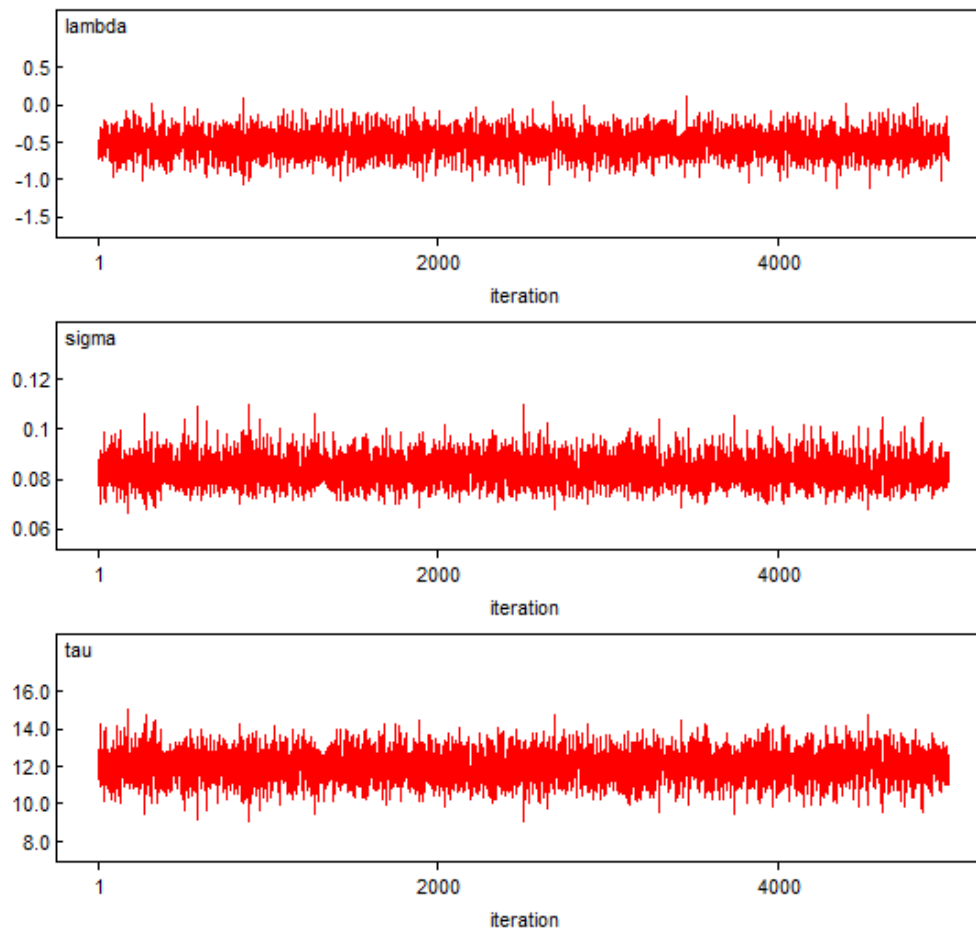


**Gambar L.2.16.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0132

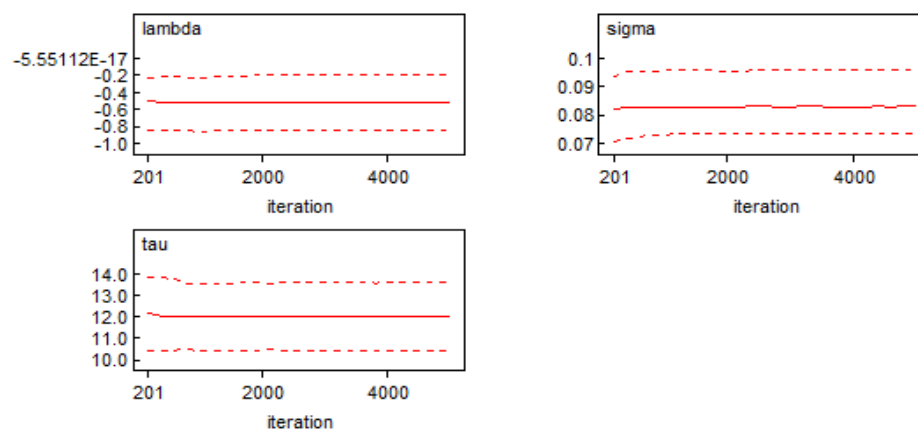


**Gambar L.2.17.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0132

### Lampiran 3. (Lanjutan)



**Gambar L.2.18.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0132



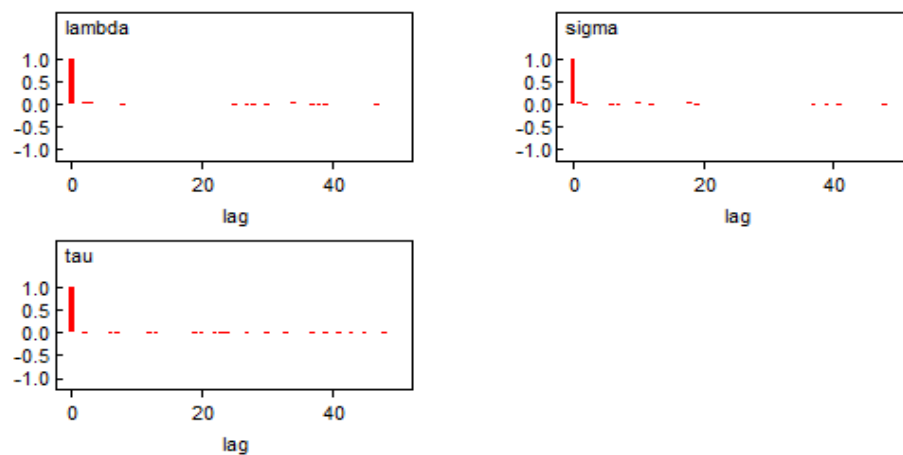
**Gambar L.2.19.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0132

### Lampiran 3. (Lanjutan)

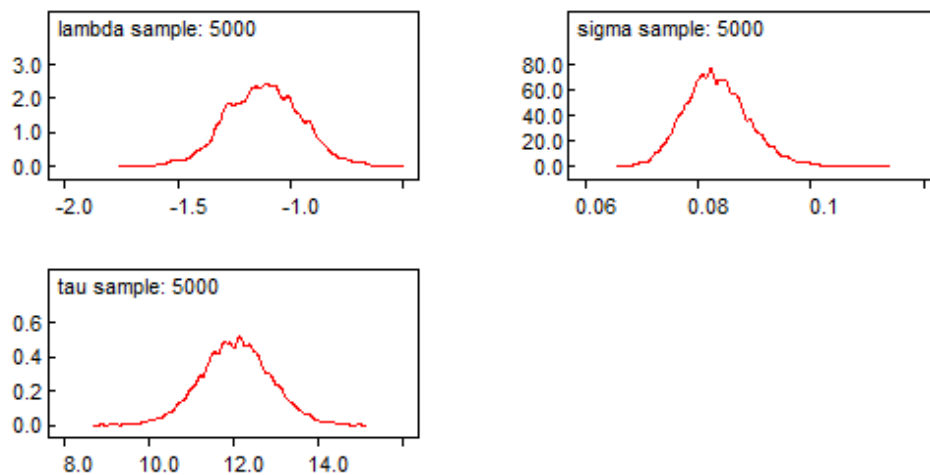
#### 6. Rata-Rata Data Simulasi untuk ID Gen LS0341 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.2.6.** Estimasi Parameter Rata-Rata Data Simulasi untuk ID Gen LS0341

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-1,116	0,1651	0,0022	-1,434	-1,116	-0,792	1	5.000
sigma	0,0834	0,0057	7,7E-05	0,0732	0,0829	0,0958	1	5.000
tau	12,05	0,8181	0,0108	10,44	12,06	13,67	1	5.000

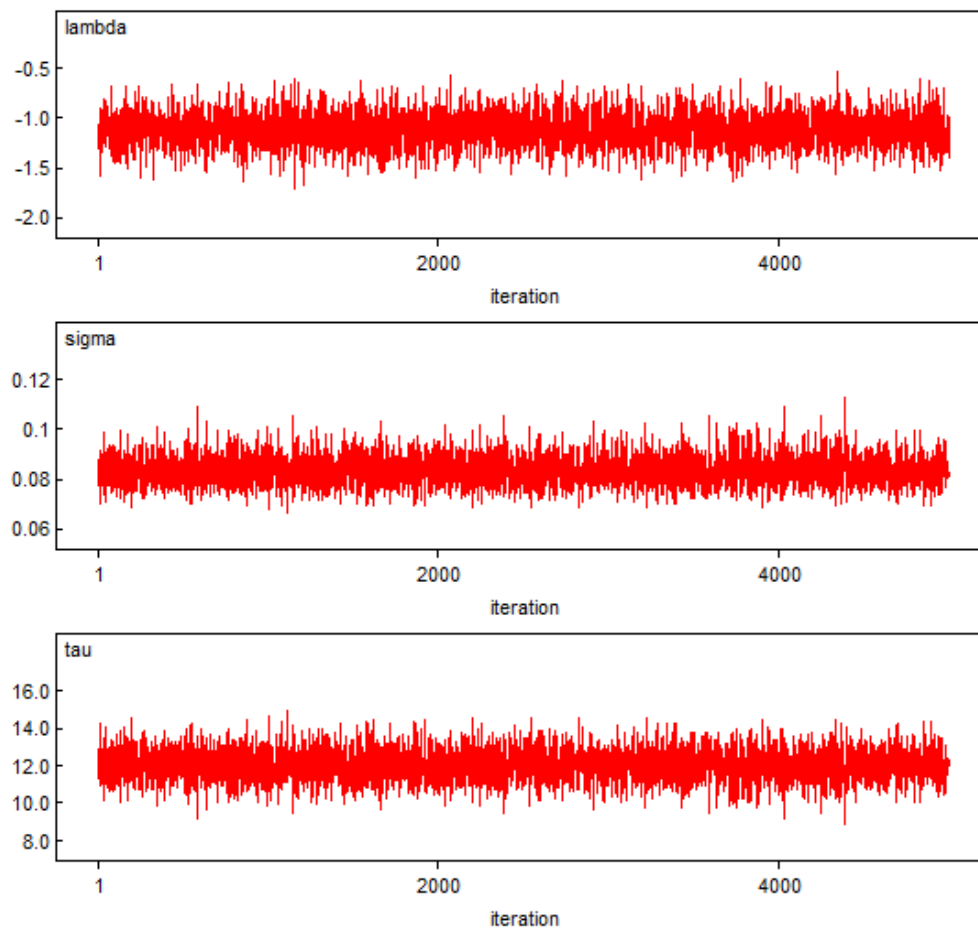


**Gambar L.2.20.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0341

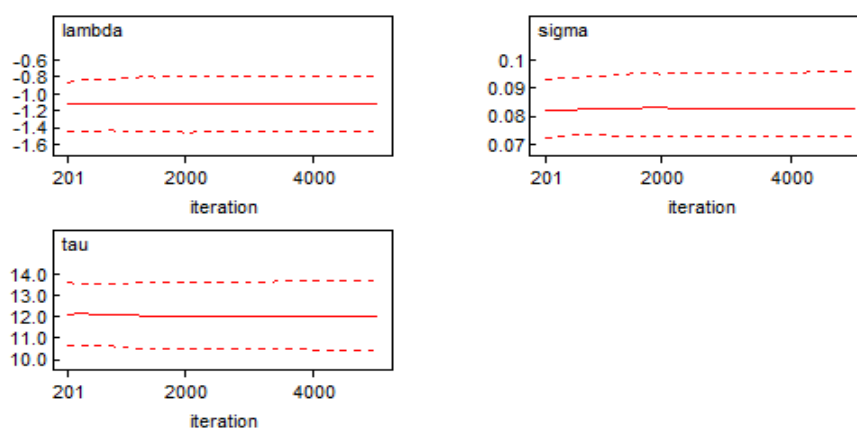


**Gambar L.2.21.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0341

### Lampiran 3. (Lanjutan)



**Gambar L.2.22.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0341



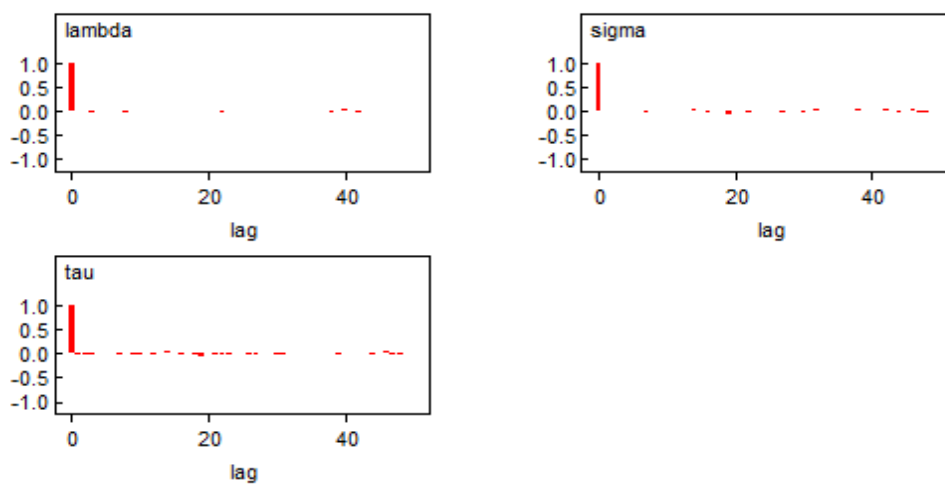
**Gambar L.2.23.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0341

**Lampiran 4.** Hasil *Running* Program BUGS untuk 7 ID Gen pada model BUM-Normal Data *Chickpea*

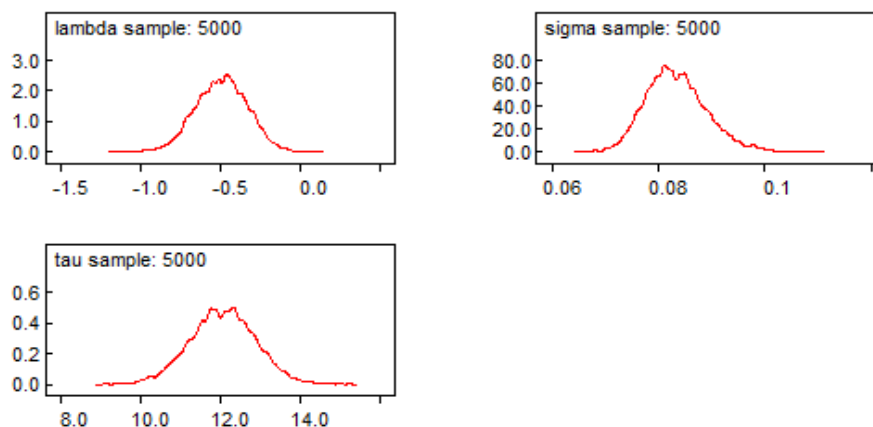
1. ID Gen LS0612 (Pembangkitan Parameter 5.000 kali)

**Tabel L.3.1.** Estimasi Parameter Model BUM-Normal untuk ID Gen LS0612

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,4944	0,1645	0,0025	-0,8167	-0,491	-0,1798	1	5.000
sigma	0,0834	0,0058	7,2E-02	0,0733	0,0829	0,0959	1	5.000
tau	12,05	0,8169	0,0102	10,43	12,06	13,64	1	5.000

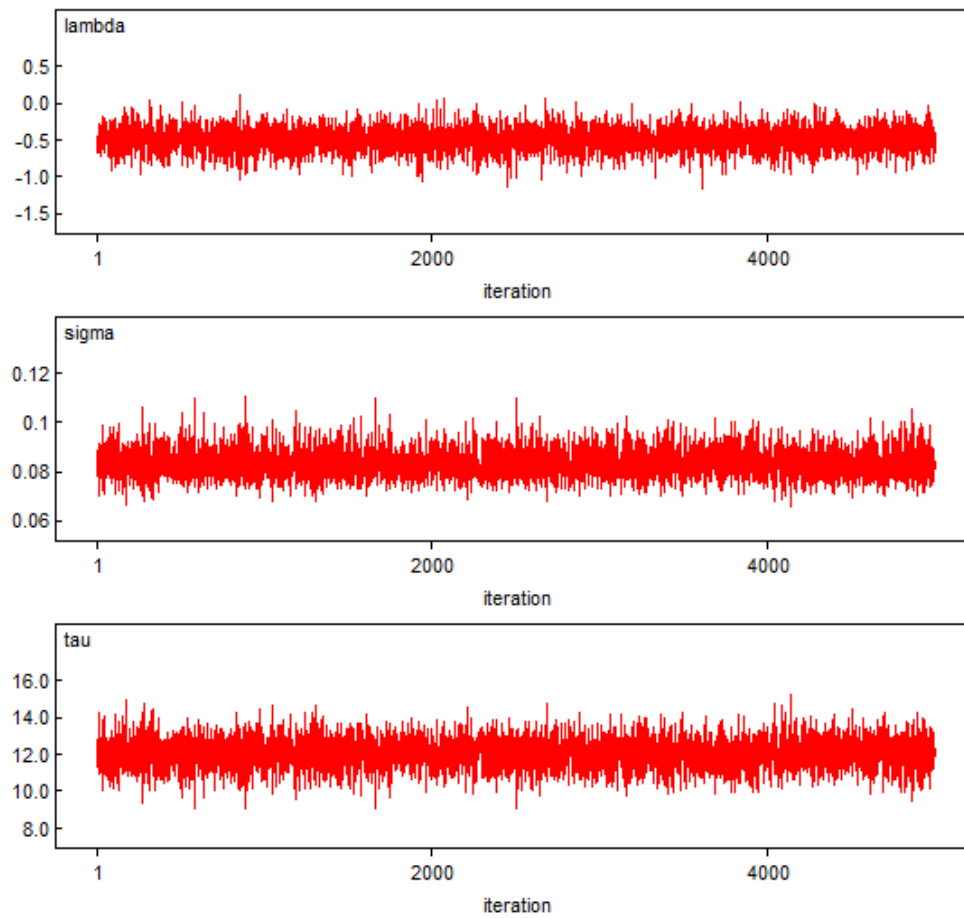


**Gambar L.3.1.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0612

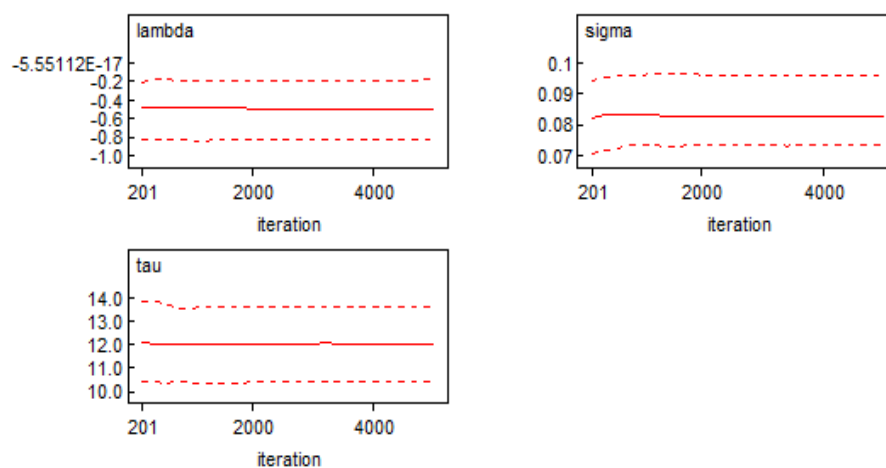


**Gambar L.3.2.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0612

#### Lampiran 4. (Lanjutan)



**Gambar L.3.3.** *History* Data Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0612



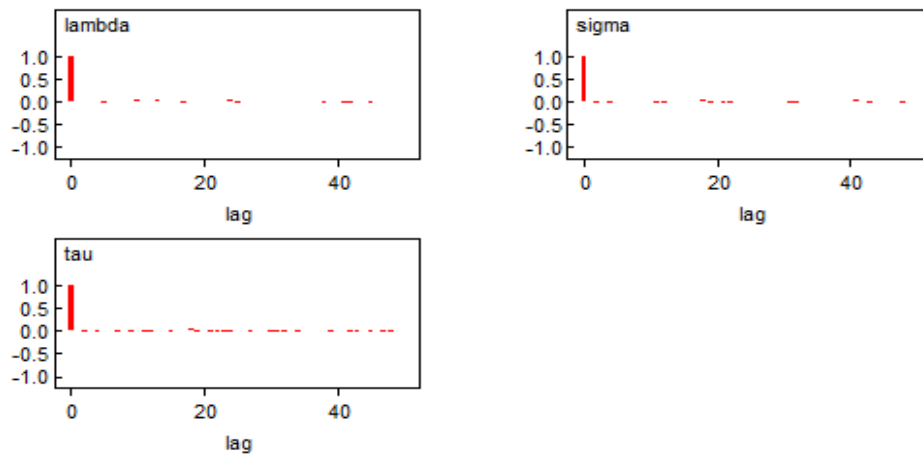
**Gambar L.3.4.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0612

#### Lampiran 4. (Lanjutan)

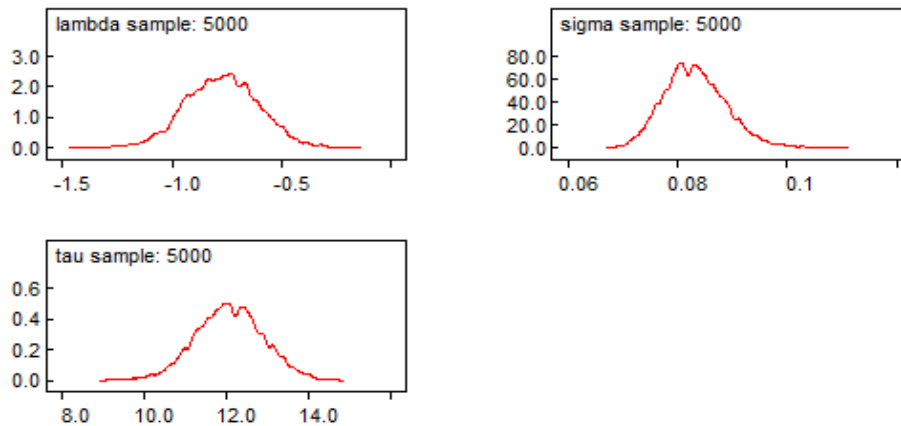
##### 2. ID Gen LS0759 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.3.2.** Estimasi Parameter Model BUM-Normal untuk ID Gen LS0759

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,7703	0,1678	0,0026	-1,098	-0,7689	-0,4426	1	5.000
sigma	0,0833	0,0057	7,3E-02	0,0732	0,0829	0,0956	1	5.000
tau	12,07	0,8144	0,0101	10,47	12,06	13,66	1	5.000



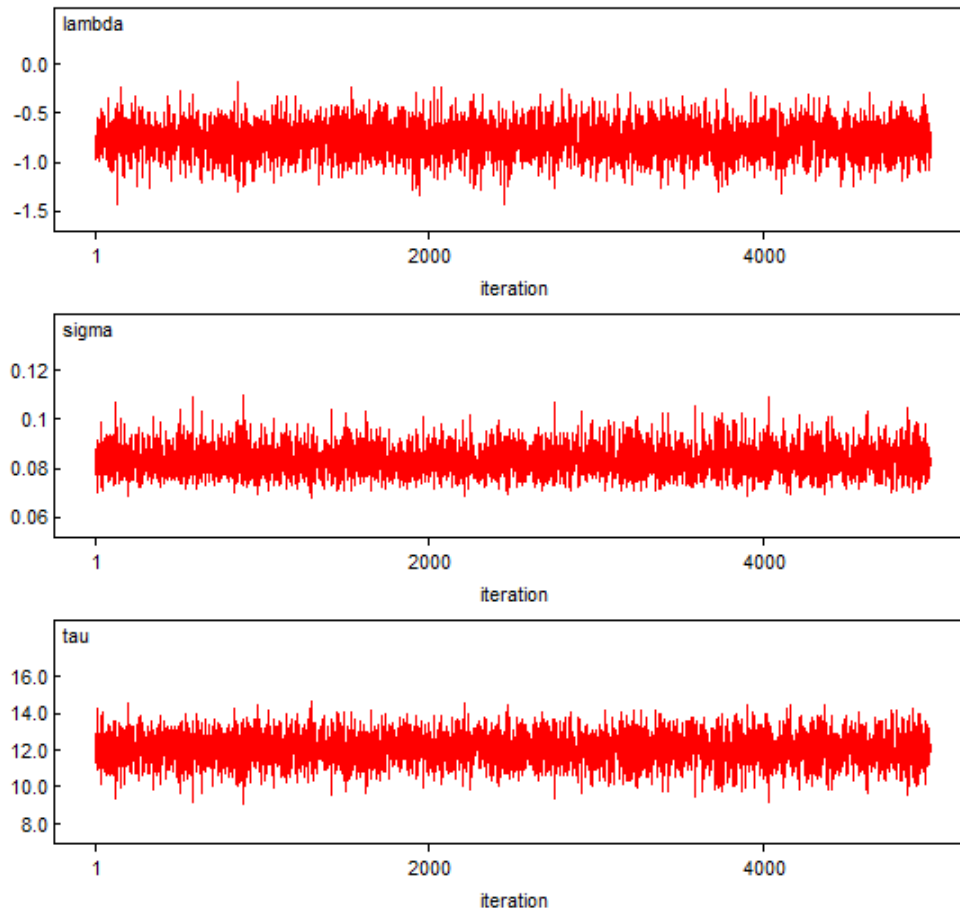
**Gambar L.3.5.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0759



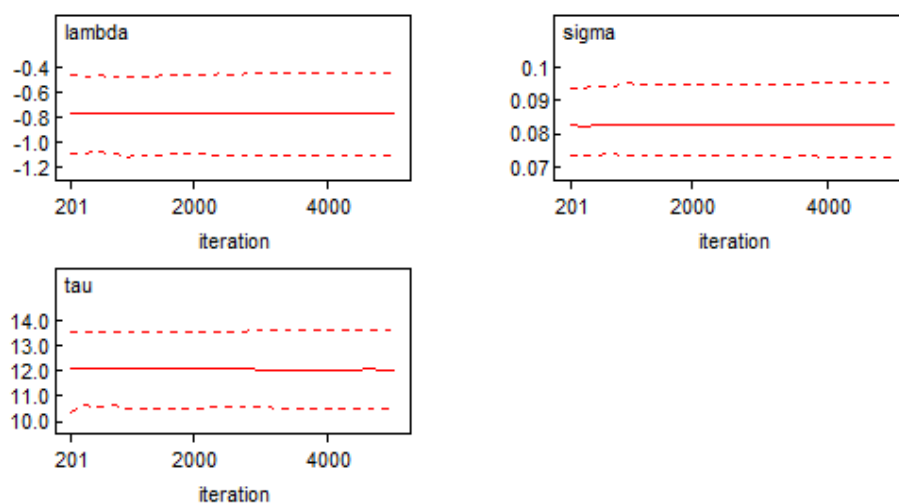
**Gambar L.3.6.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0759



#### Lampiran 4. (Lanjutan)



**Gambar L.3.7.** *History* Data Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0759



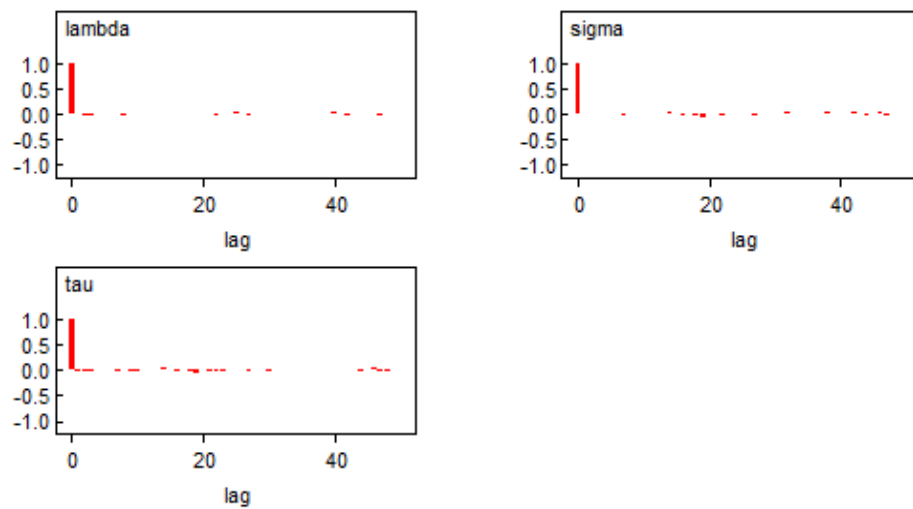
**Gambar L.3.8.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0759

#### Lampiran 4. (Lanjutan)

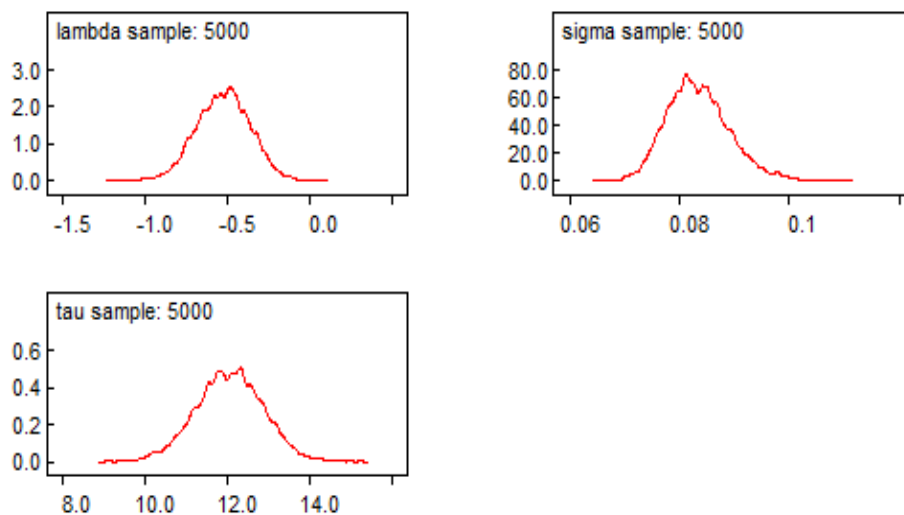
##### 3. ID Gen U278 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.3.3.** Estimasi Parameter Model BUM-Normal untuk ID Gen U278

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,5254	0,1643	0,0024	-0,8466	-0,5221	-0,2121	1	5.000
sigma	0,0834	0,0057	7,4E-05	0,0735	0,0829	0,09596	1	5.000
tau	12,05	0,8151	0,0105	10,42	12,06	13,62	1	5.000

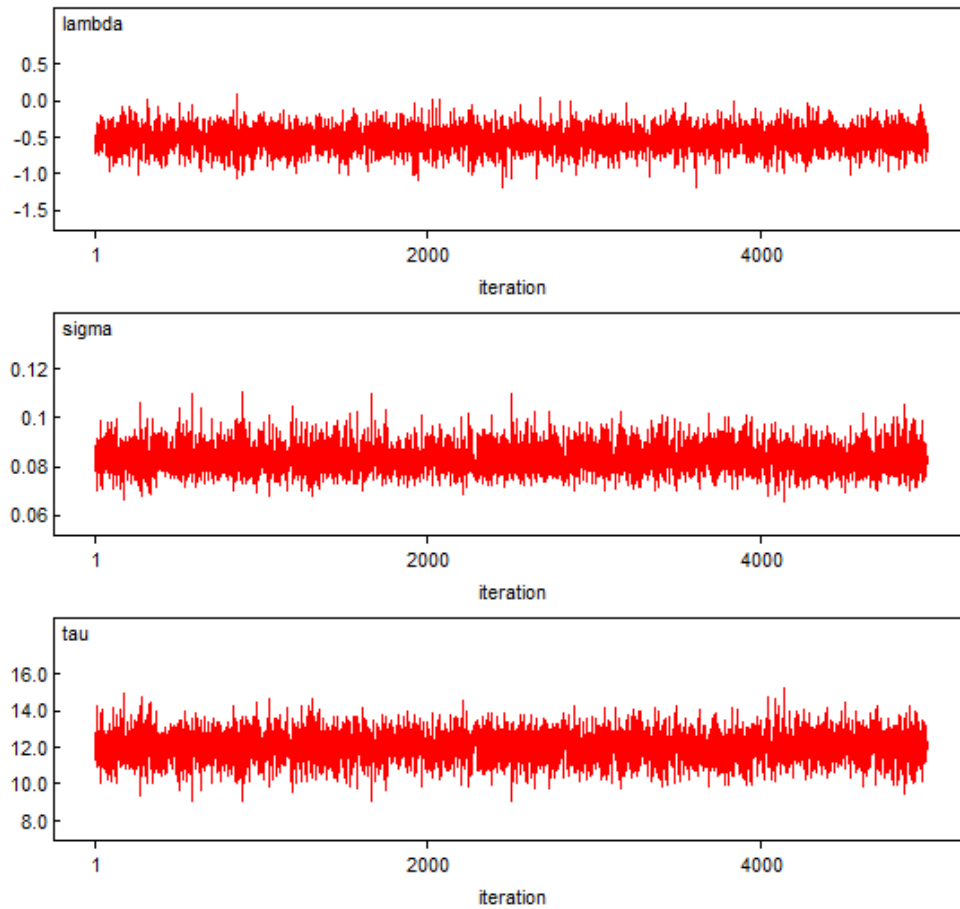


**Gambar L.3.9.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen U278

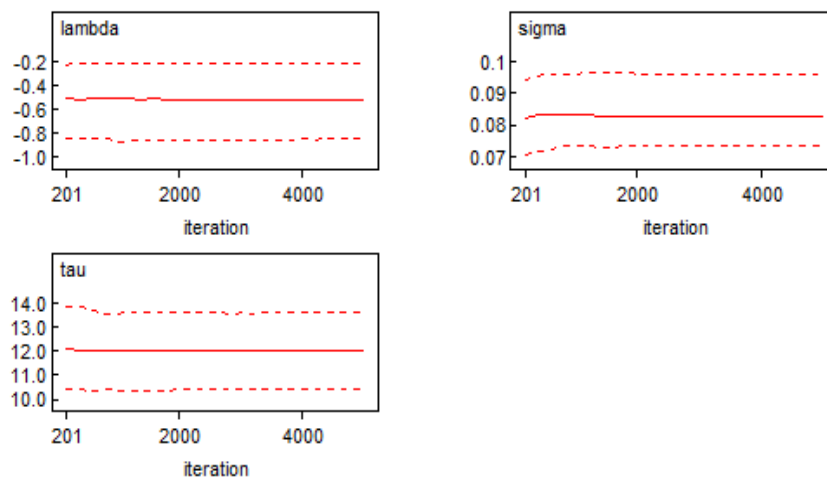


**Gambar L.3.10.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen U278

#### Lampiran 4. (Lanjutan)



**Gambar L.3.11.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen U278



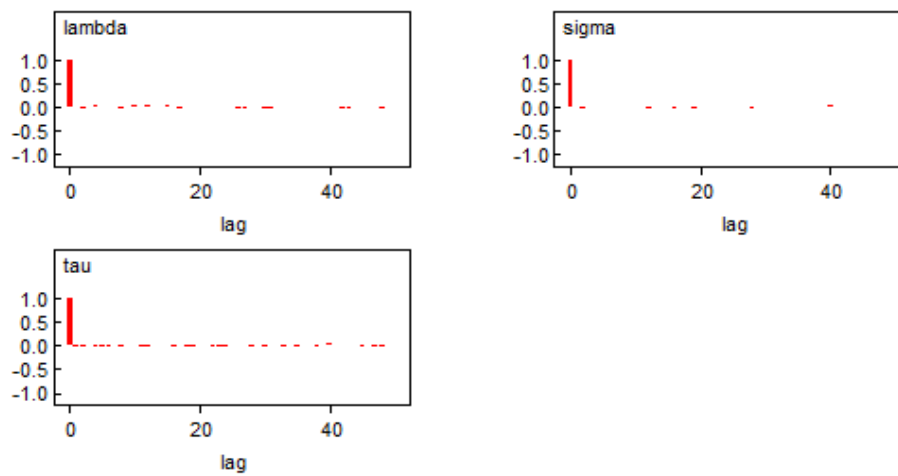
**Gambar L.3.12.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen U278

#### Lampiran 4. (Lanjutan)

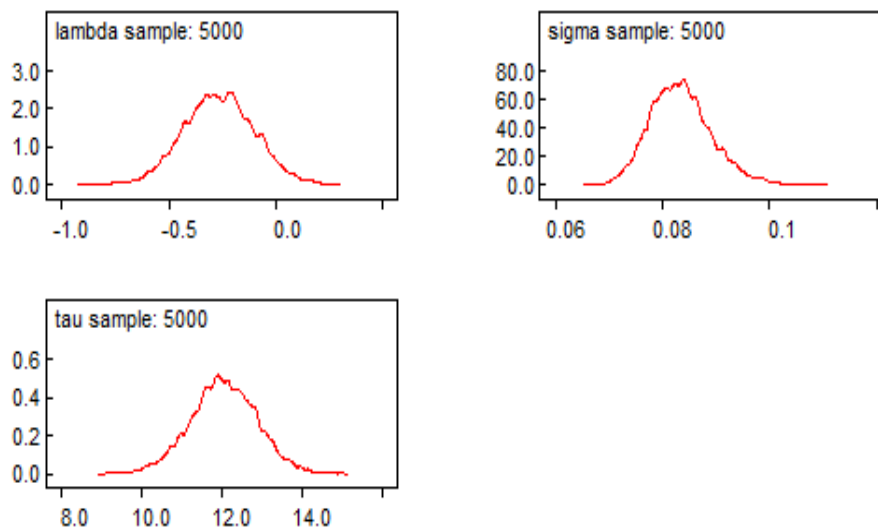
#### 4. ID Gen U017 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.3.4.** Estimasi Parameter Model BUM-Normal untuk ID Gen U017

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,2641	0,1661	0,0022	-0,5871	-0,2647	0,0667	1	5.000
sigma	0,0834	0,0058	7,3E-02	0,0732	0,0831	0,0961	1	5.000
tau	12,04	0,8186	0,0101	10,4	12,04	13,65	1	5.000

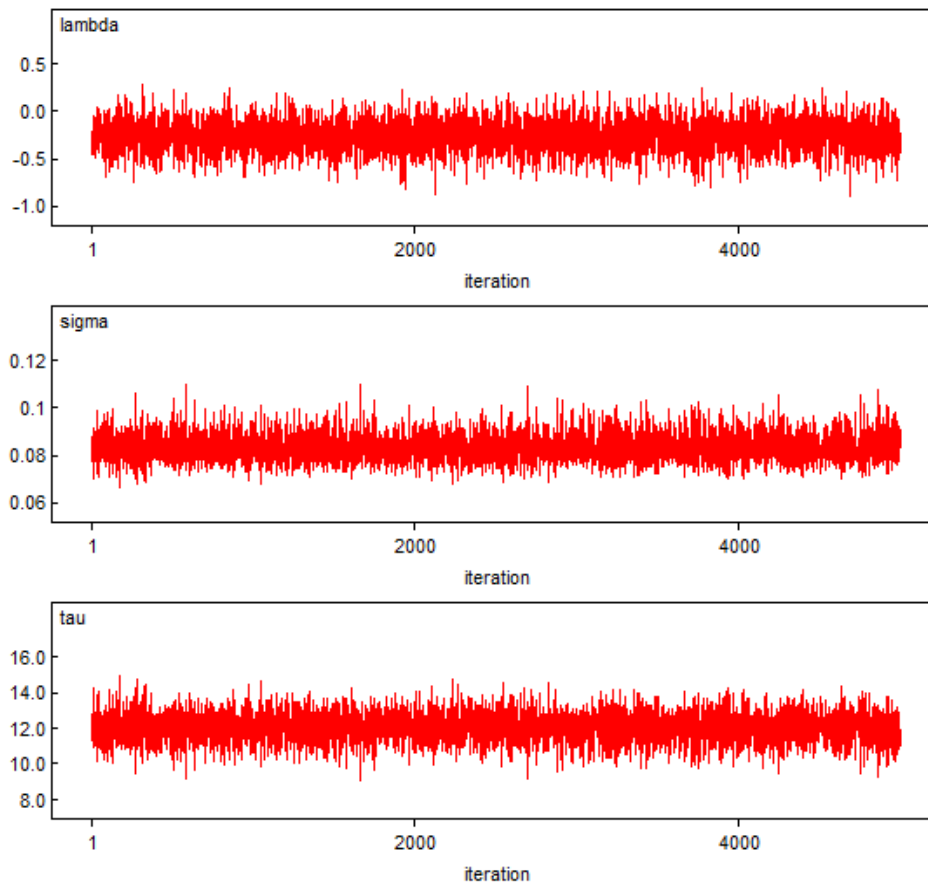


**Gambar L.3.13.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen U017

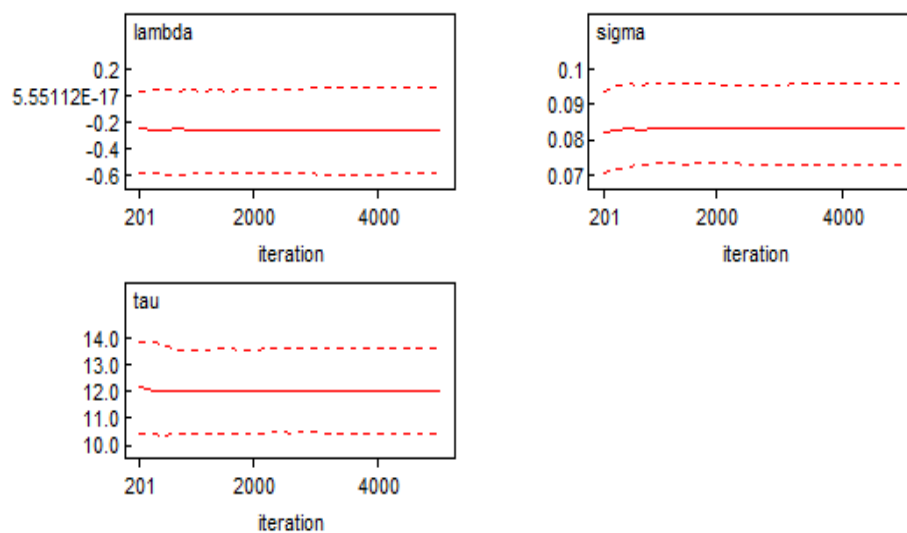


**Gambar L.3.14.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen U017

#### Lampiran 4. (Lanjutan)



**Gambar L.3.15.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen U017



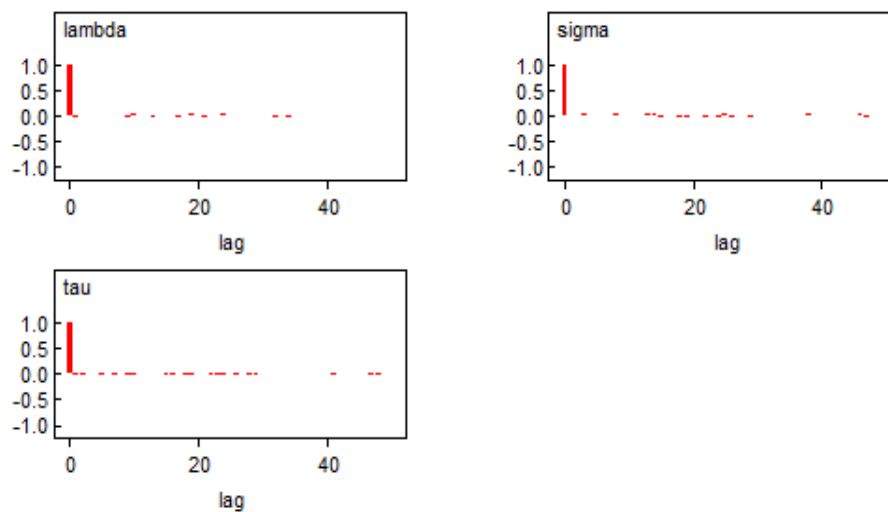
**Gambar L.3.16.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen U017

#### Lampiran 4. (Lanjutan)

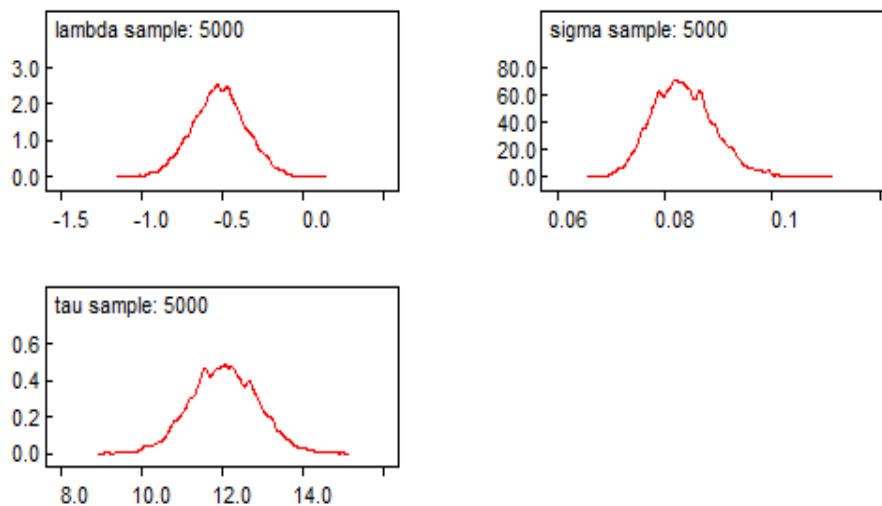
##### 5. ID Gen LS0132 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.3.5.** Estimasi Parameter Model BUM-Normal untuk ID Gen LS0132

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-0,5207	0,1667	0,0025	-0,8448	-0,5205	-0,194	1	5.000
sigma	0,0835	0,0058	7,7E-02	0,0734	0,0831	0,0957	1	5.000
tau	12,04	0,8183	0,0109	10,45	12,03	13,63	1	5.000

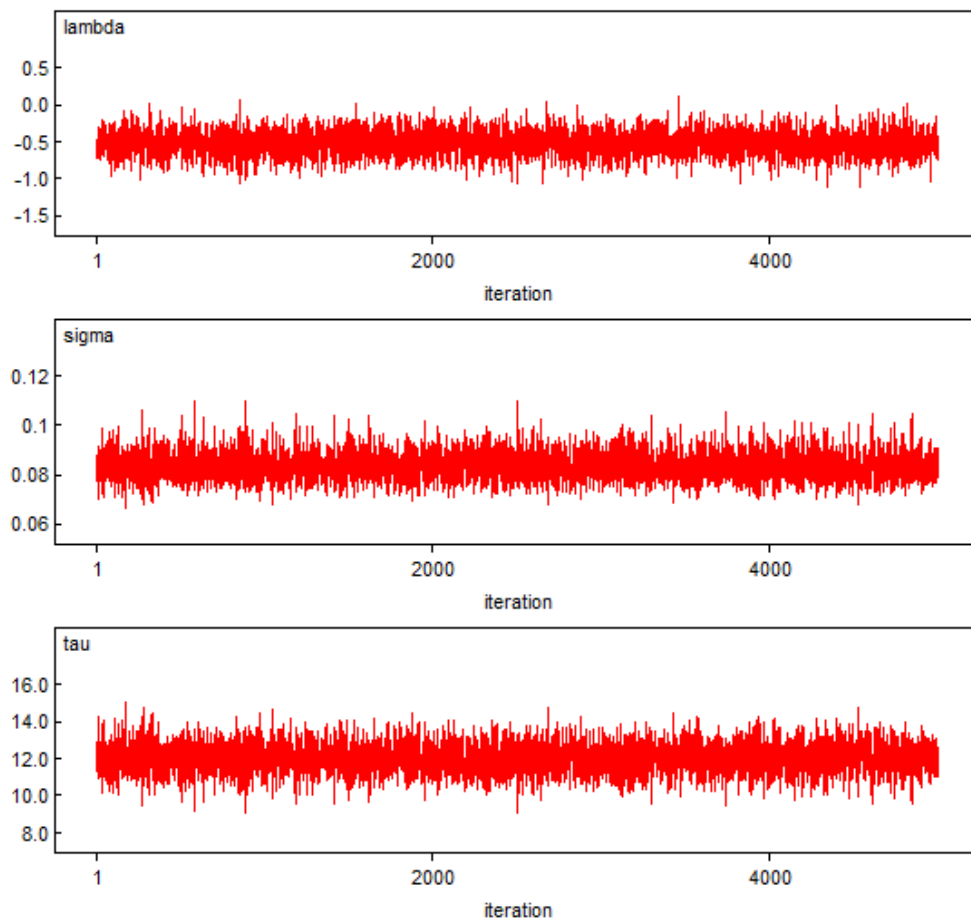


**Gambar L.3.17.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0132

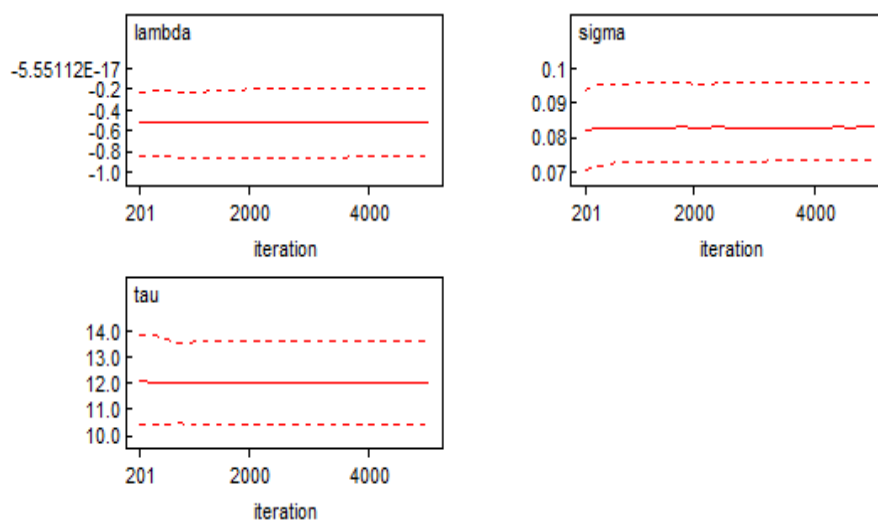


**Gambar L.3.18.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0132

#### Lampiran 4. (Lanjutan)



**Gambar L.3.19.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0132



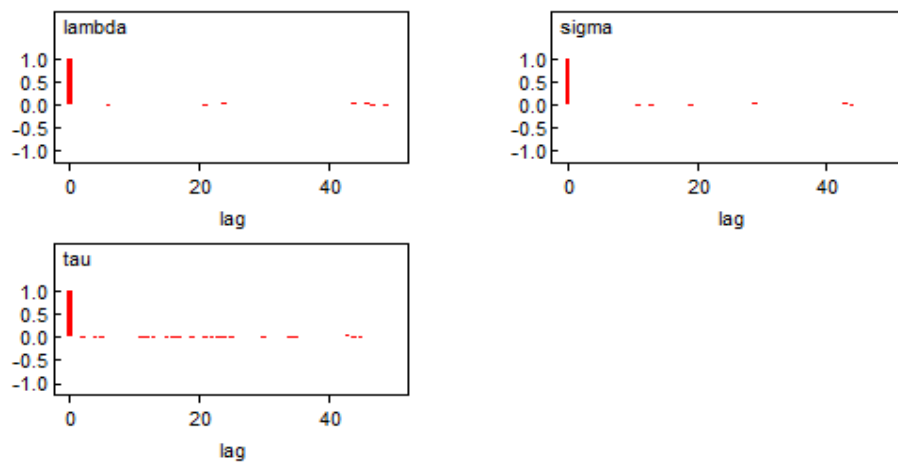
**Gambar L.3.20.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0132

#### Lampiran 4. (Lanjutan)

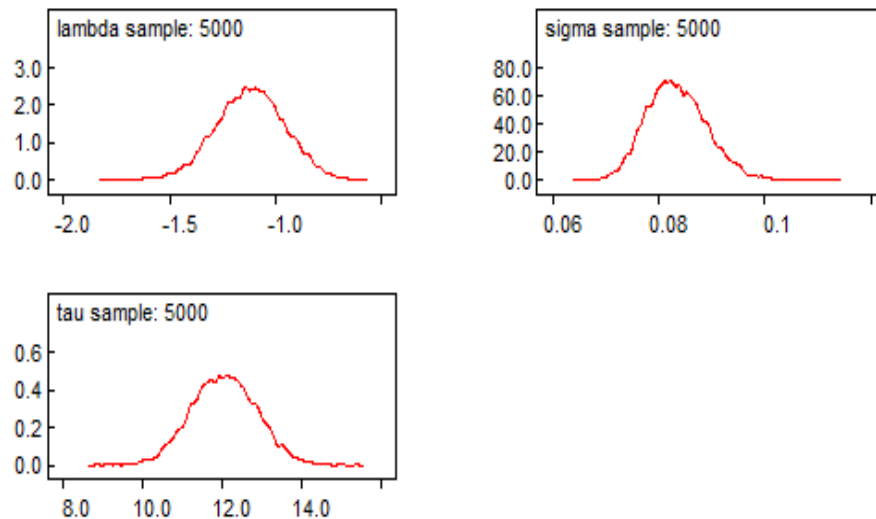
6. ID Gen LS0341 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.3.6.** Estimasi Parameter Model BUM-Normal untuk ID Gen LS0341

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda	-1,117	0,1636	0,0024	-1,441	-1,115	-0,8018	1	5.000
sigma	0,0834 1	0,0057	7,1E-02	0,0733	0,0830	0,0953	1	5.000
tau	12,04	0,8114	0,0101	10,49	12,04	13,64	1	5.000



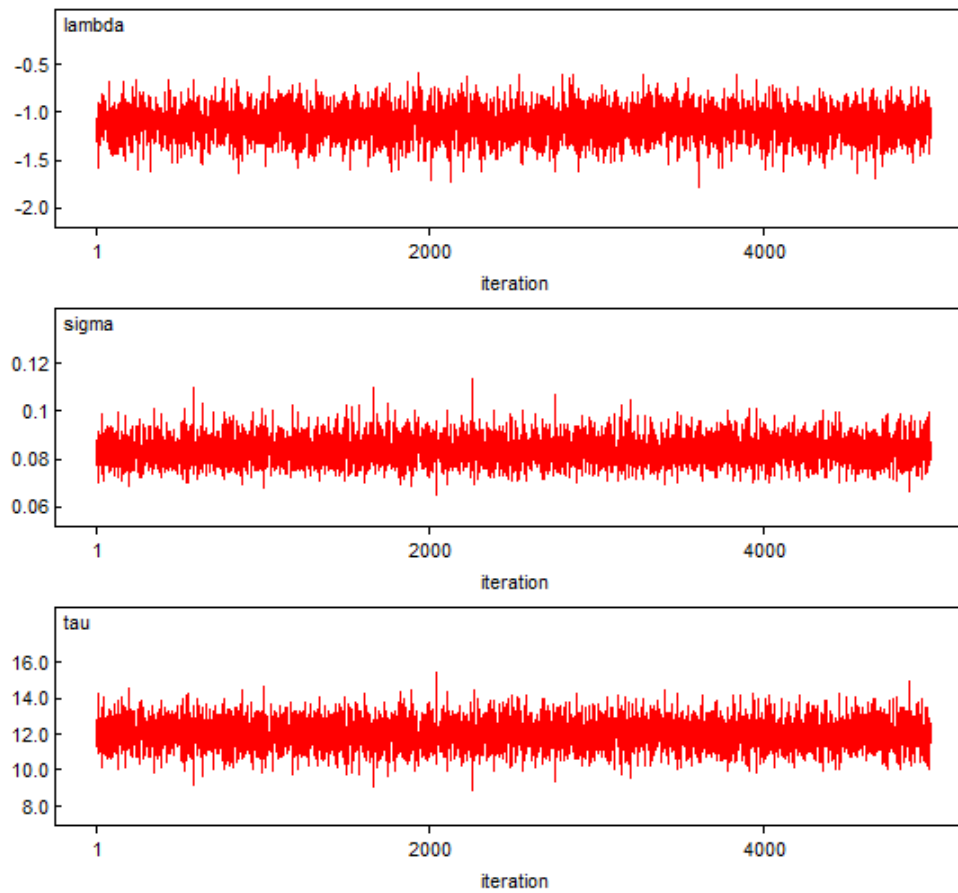
**Gambar L.3.21.** Fungsi Autokorelasi setiap Parameter Model BUM-Normal untuk ID Gen LS0341



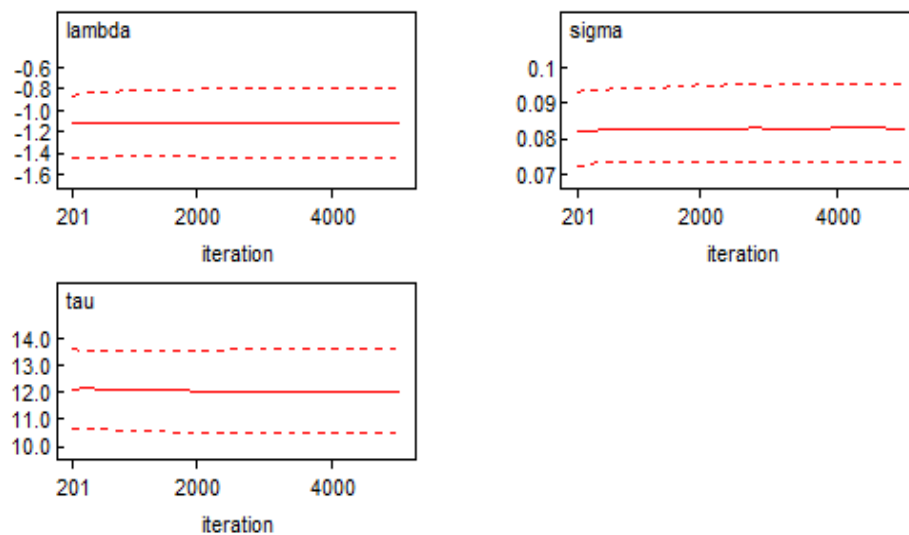
**Gambar L.3.22.** Densitas Kernel setiap Parameter Model BUM-Normal untuk ID Gen LS0341



#### Lampiran 4. (Lanjutan)



**Gambar L.3.23.** *History Data* Bangkitan setiap Parameter Model BUM-Normal untuk ID Gen LS0341



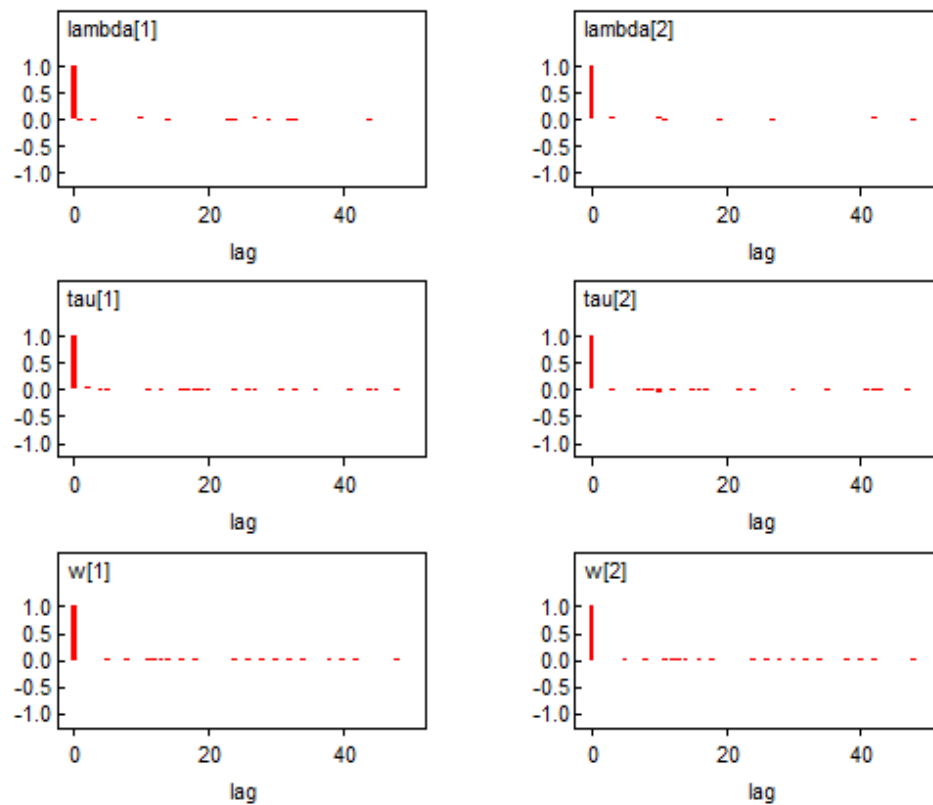
**Gambar L.3.24.** *Quantiles* setiap Parameter Model BUM-Normal untuk ID Gen LS0341

**Lampiran 5.** Hasil *Running* Program BUGS untuk 19 ID Gen pada Model BMM-Normal Data Simulasi

1. ID Gen LS0616 (Pembangkitan Parameter 5.000 Kali)

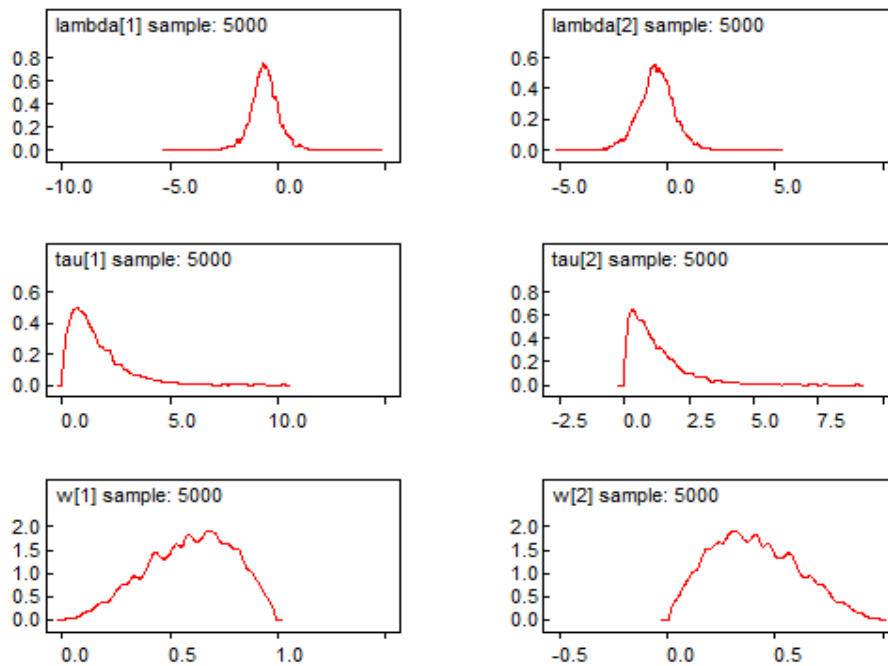
**Tabel L.4.1.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0616

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5872	0,6511	0,0081	-1,894	-0,6026	0,7304	1	5.000
lambda[2]	-0,5121	0,8272	0,0111	-2,193	-0,505	1,125	1	5.000
tau[1]	1,506	1,163	0,0166	0,1495	1,204	4,525	1	5.000
tau[2]	1,207	1,042	0,0135	0,0830	0,9153	4,062	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

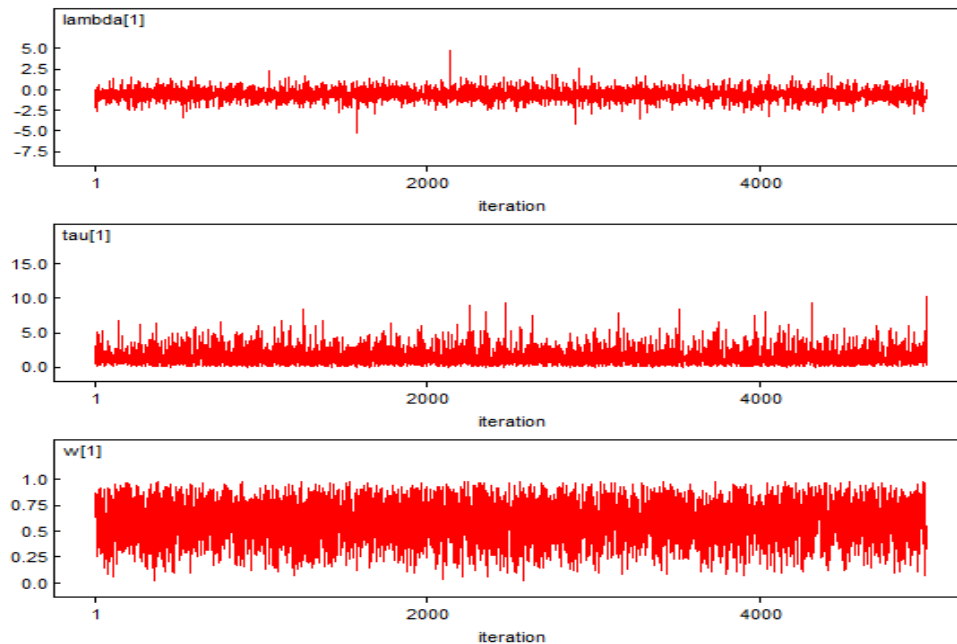


**Gambar L.4.1.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616

## Lampiran 5. (Lanjutan)

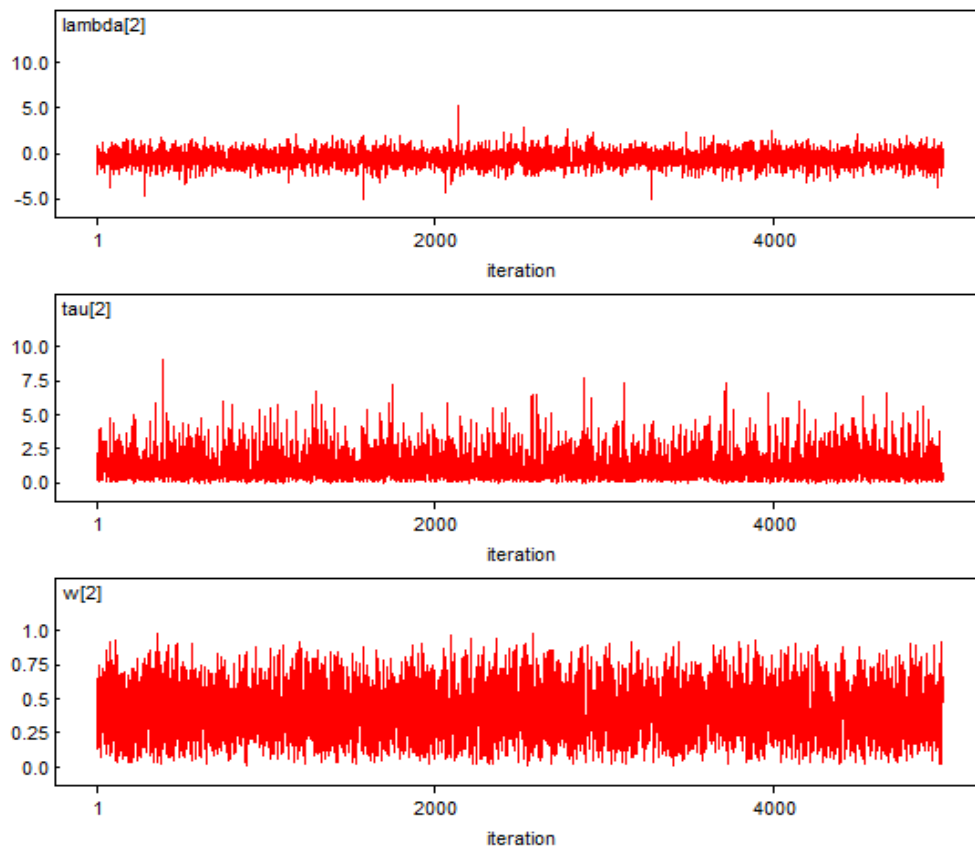


**Gambar L.4.2.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616

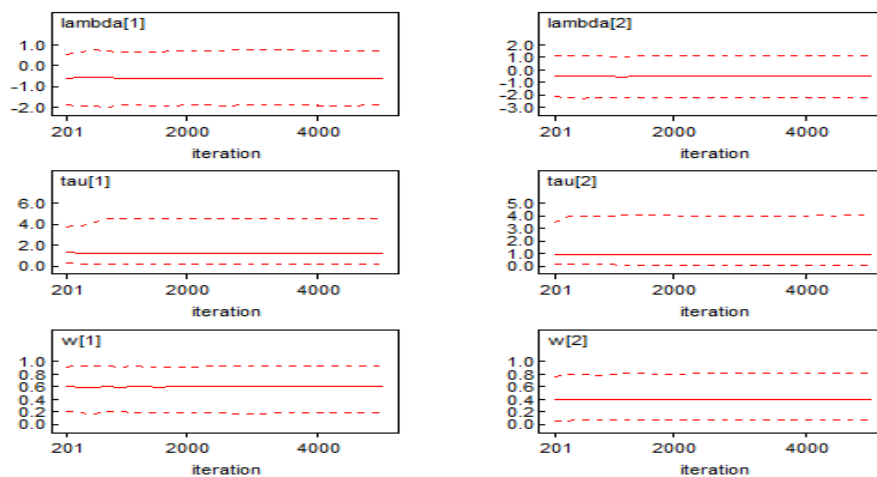


**Gambar L.4.3.1.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616

## Lampiran 5. (Lanjutan)



**Gambar L.4.3.2.** *History Data* Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616



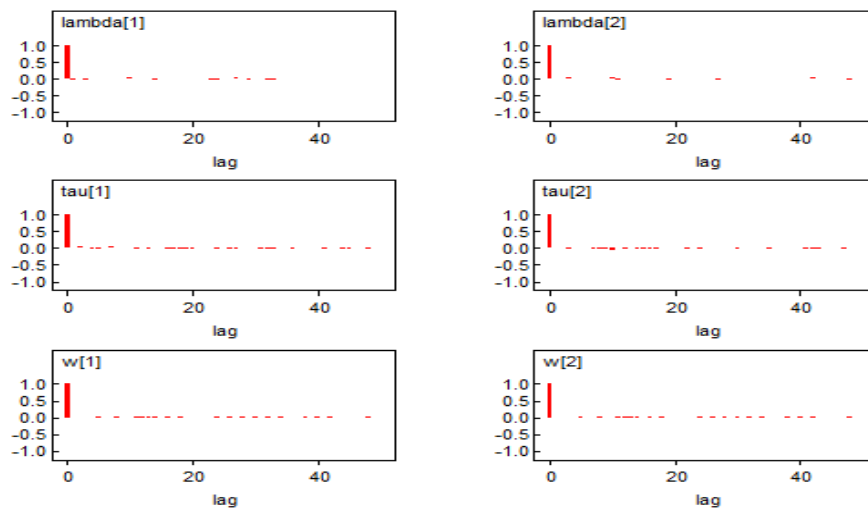
**Gambar L.4.4.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616

## Lampiran 5. (Lanjutan)

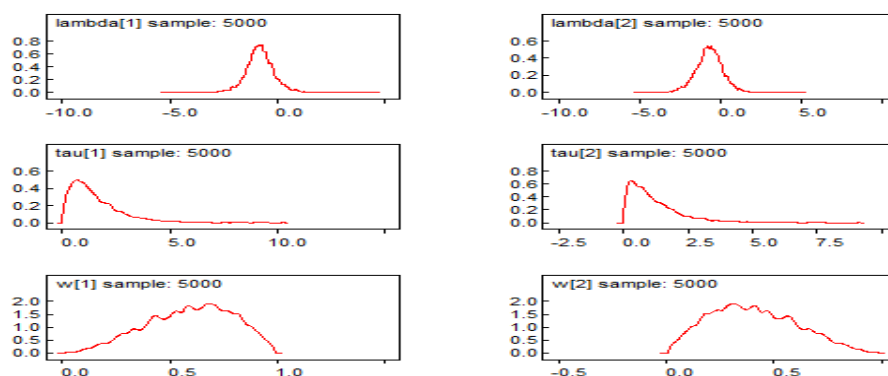
### 2. ID Gen LS0159 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.4.2.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0159

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,7995	0,6517	0,0080	-2,092	-0,821	0,5573	1	5.000
lambda[2]	-0,6761	0,8301	0,0111	-2,388	-0,6602	0,9458	1	5.000
tau[1]	1,518	1,175	0,0169	0,1499	1,215	4,564	1	5.000
tau[2]	1,203	1,041	0,0134	0,0822	0,9106	4,066	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

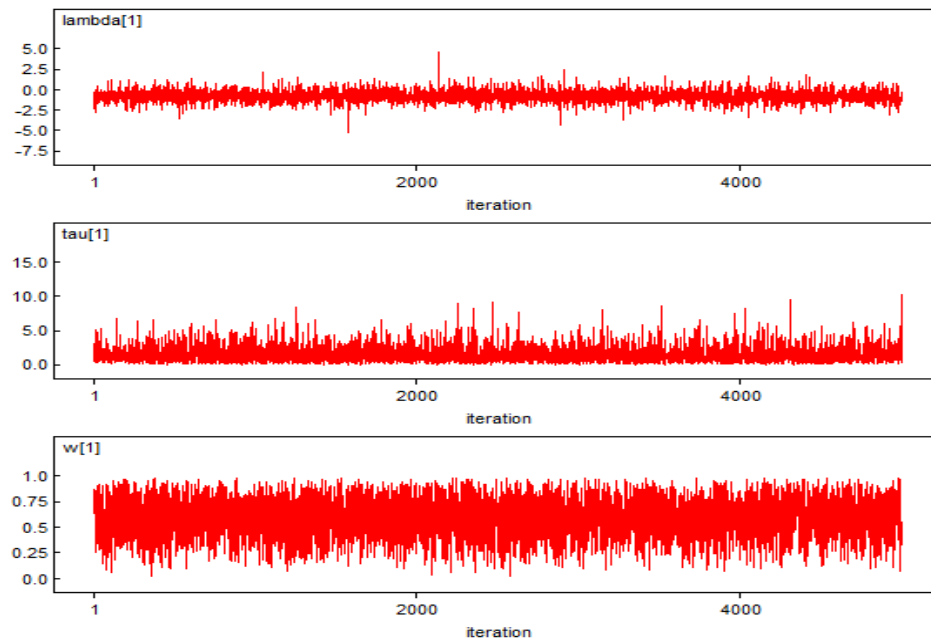


**Gambar L.4.5.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159

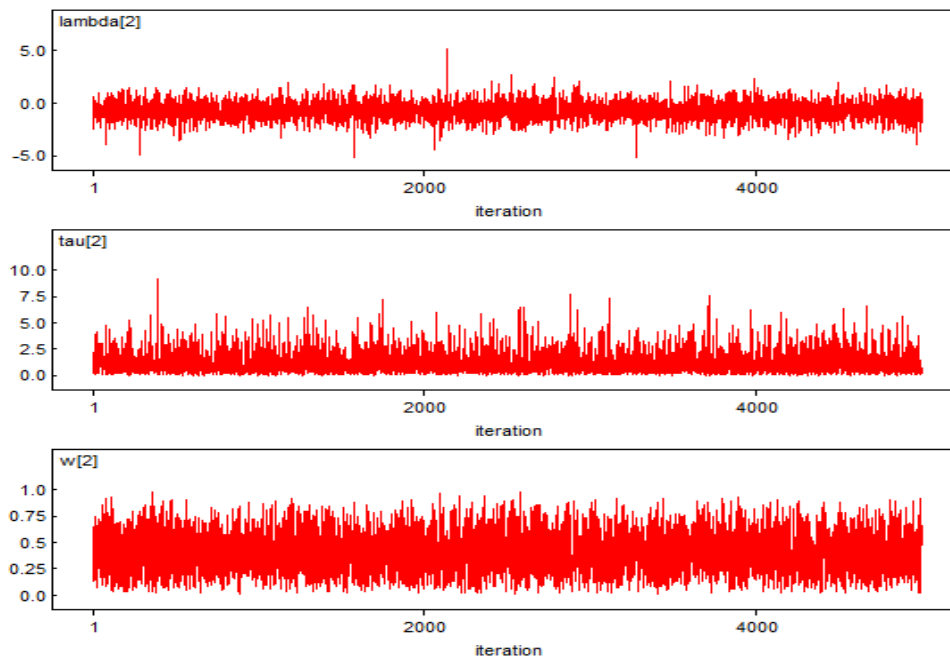


**Gambar L.4.6.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159

## Lampiran 5. (Lanjutan)

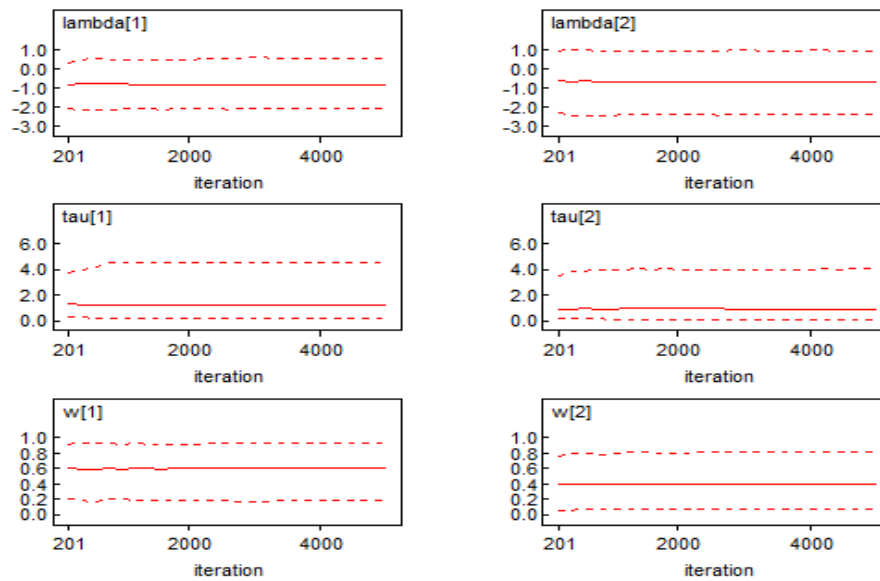


**Gambar L.4.7.1.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159*



**Gambar L.4.7.2.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159*

## Lampiran 5. (Lanjutan)

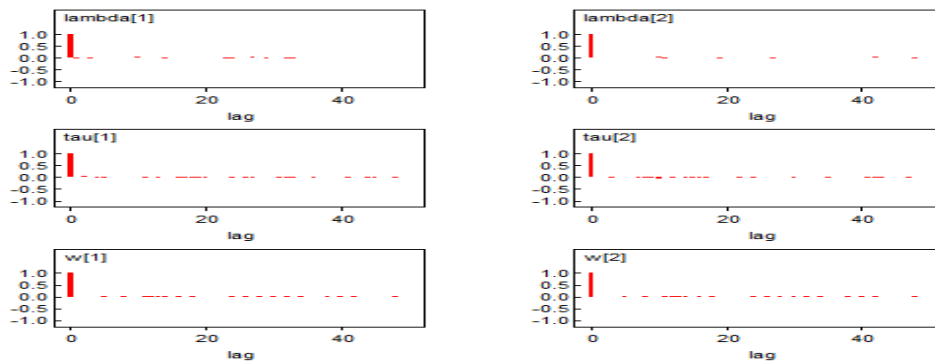


**Gambar L.4.8.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159

### 3. ID Gen LS0162 (Pembangkitan Parameter 5.000 Kali)

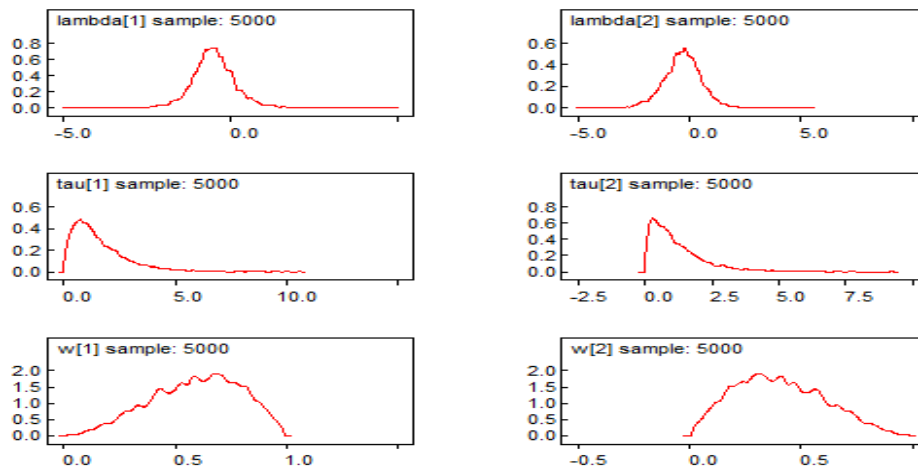
**Tabel L.4.3.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0162

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5075	0,6454	0,0079	-1,768	-0,5281	0,8622	1	5.000
lambda[2]	-0,3205	0,8337	0,0111	-2,065	-0,2981	1,288	1	5.000
tau[1]	1,582	1,228	0,0178	0,1572	1,265	4,747	1	5.000
tau[2]	1,195	1,038	0,0134	0,0826	0,9022	4,017	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

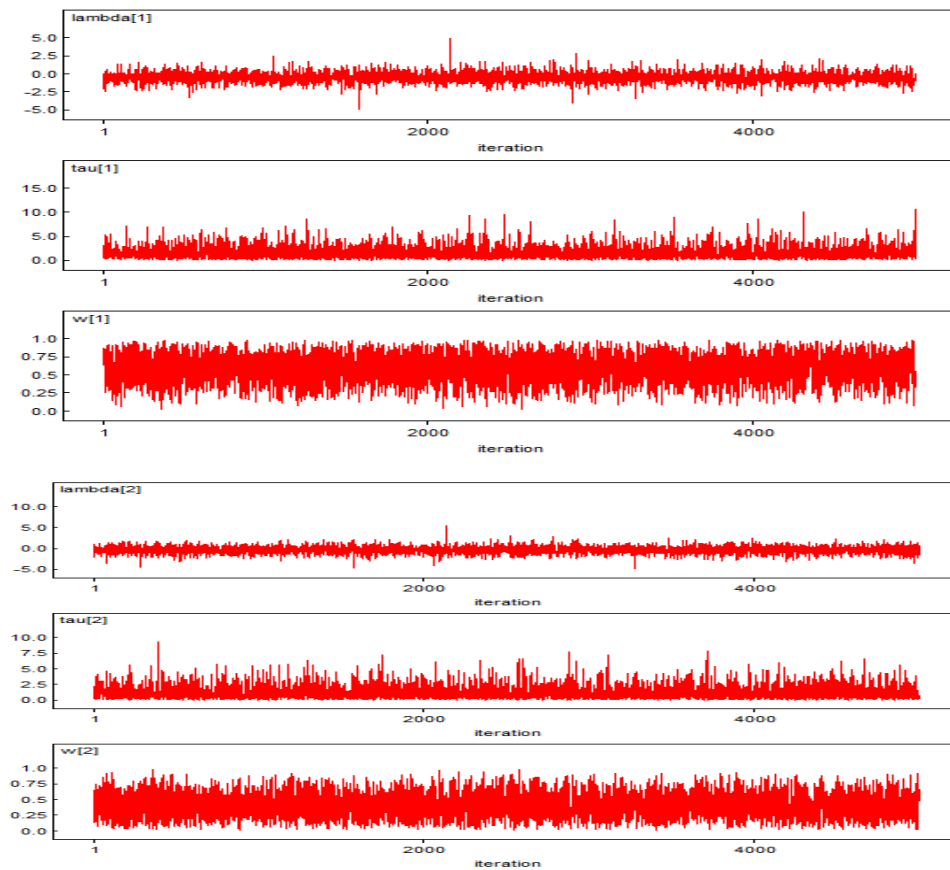


**Gambar L.4.9.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162

## Lampiran 5. (Lanjutan)



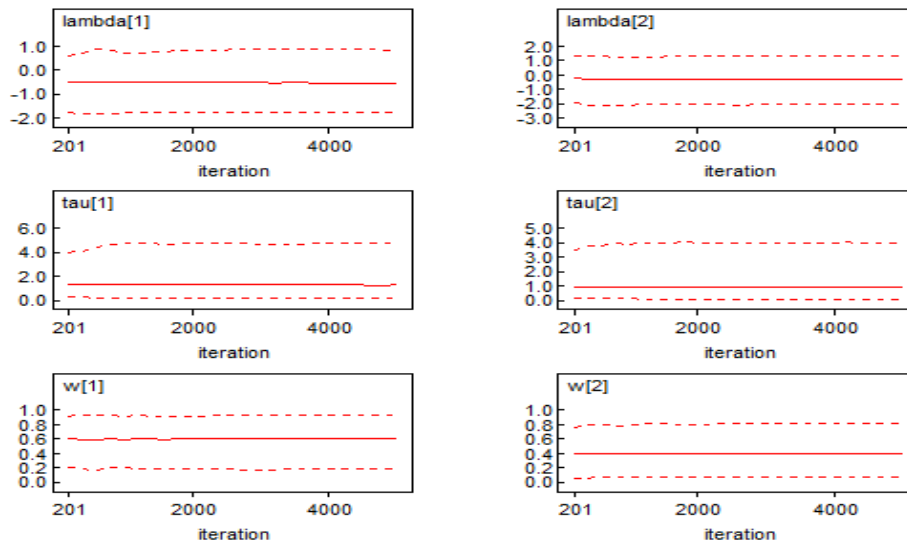
**Gambar L.4.10.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162



**Gambar L.4.11.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162



## Lampiran 5. (Lanjutan)

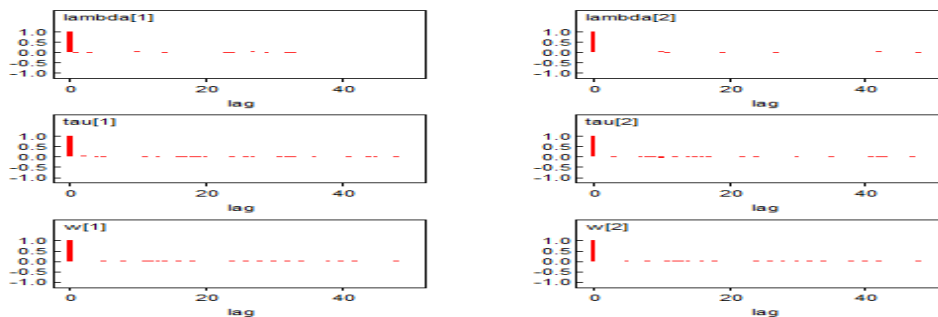


**Gambar L.4.12.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162

## 4. ID Gen LS0185 (Pembangkitan Parameter 5.000 Kali)

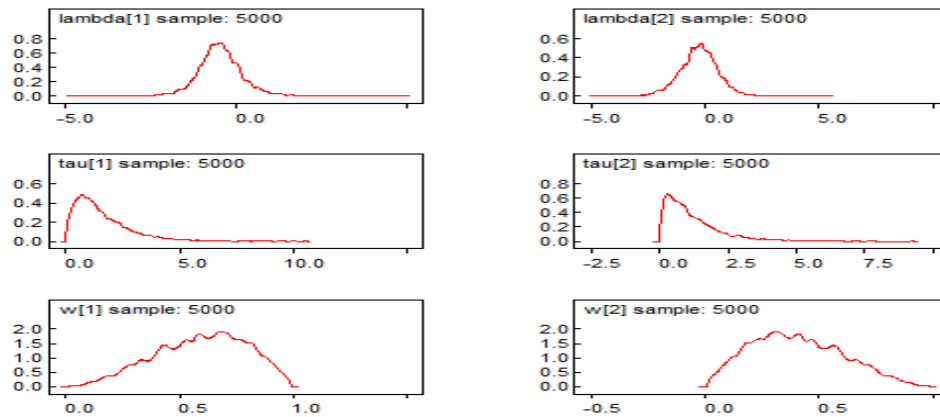
**Tabel L.4.4.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0185

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,4813	0,6464	0,0079	-1,743	-0,5022	0,8872	1	5.000
lambda[2]	-0,2856	0,8349	0,0111	-2,033	-0,262	1,328	1	5.000
tau[1]	1,579	1,226	0,0178	0,1565	1,262	4,737	1	5.000
tau[2]	1,193	1,038	0,0134	0,0821	0,9011	4,012	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

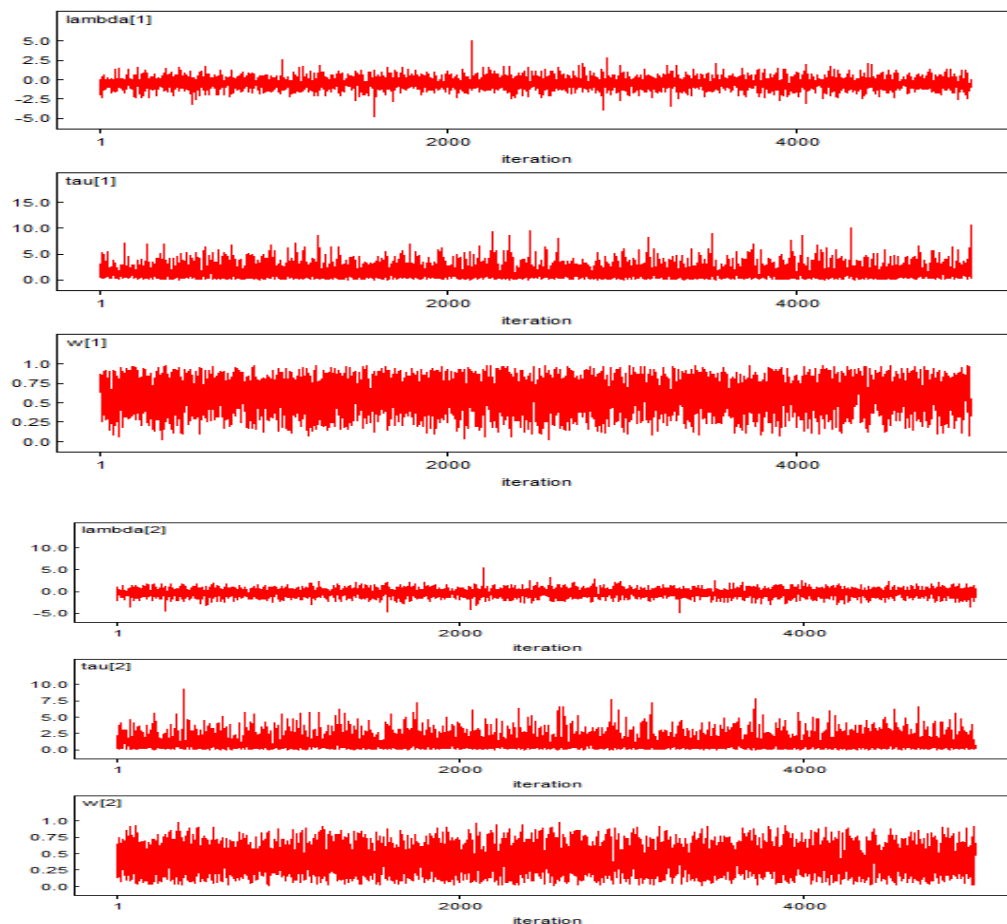


**Gambar L.4.13.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0185

## Lampiran 5. (Lanjutan)

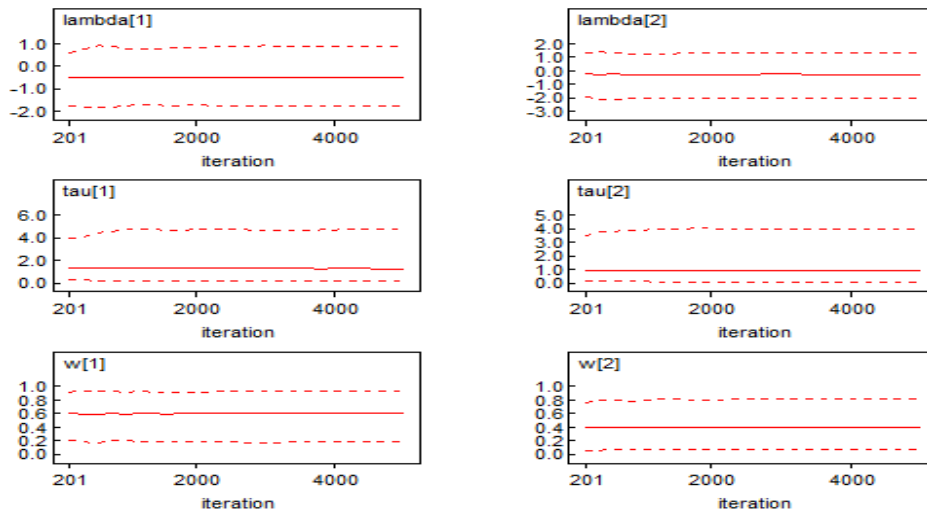


**Gambar L.4.14.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0185



**Gambar L.4.15.** *History* Data Bangkitan setiap Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0185

## Lampiran 5. (Lanjutan)

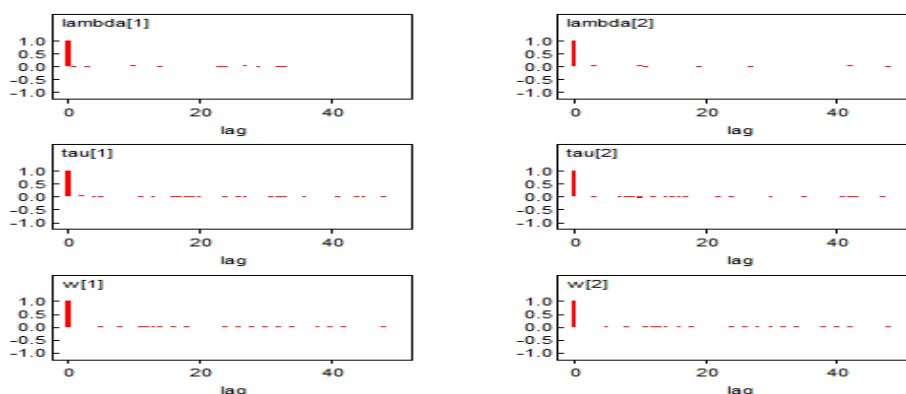


**Gambar L.4.16.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0185

## 5. ID Gen LS0081 (Pembangkitan Parameter 5.000 Kali)

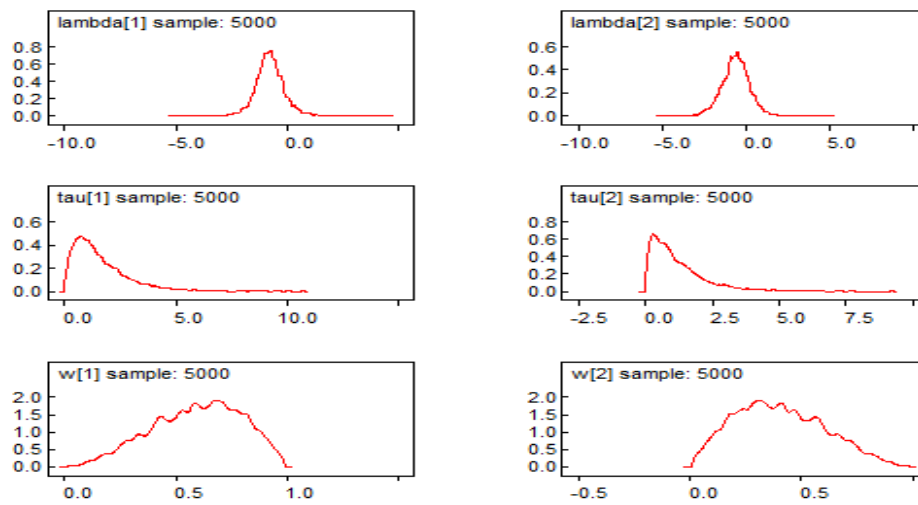
**Tabel L.4.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0081

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
$\lambda[1]$	-0,7705	0,6433	0,0079	-2,033	-0,792	0,5845	1	5.000
$\lambda[2]$	-0,6042	0,831	0,0111	-2,335	-0,5833	1,005	1	5.000
$\tau[1]$	1,593	1,237	0,0179	0,1584	1,274	4,782	1	5.000
$\tau[2]$	1,198	1,04	0,0134	0,0824	0,9043	4,039	1	5.000
$w[1]$	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
$w[2]$	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

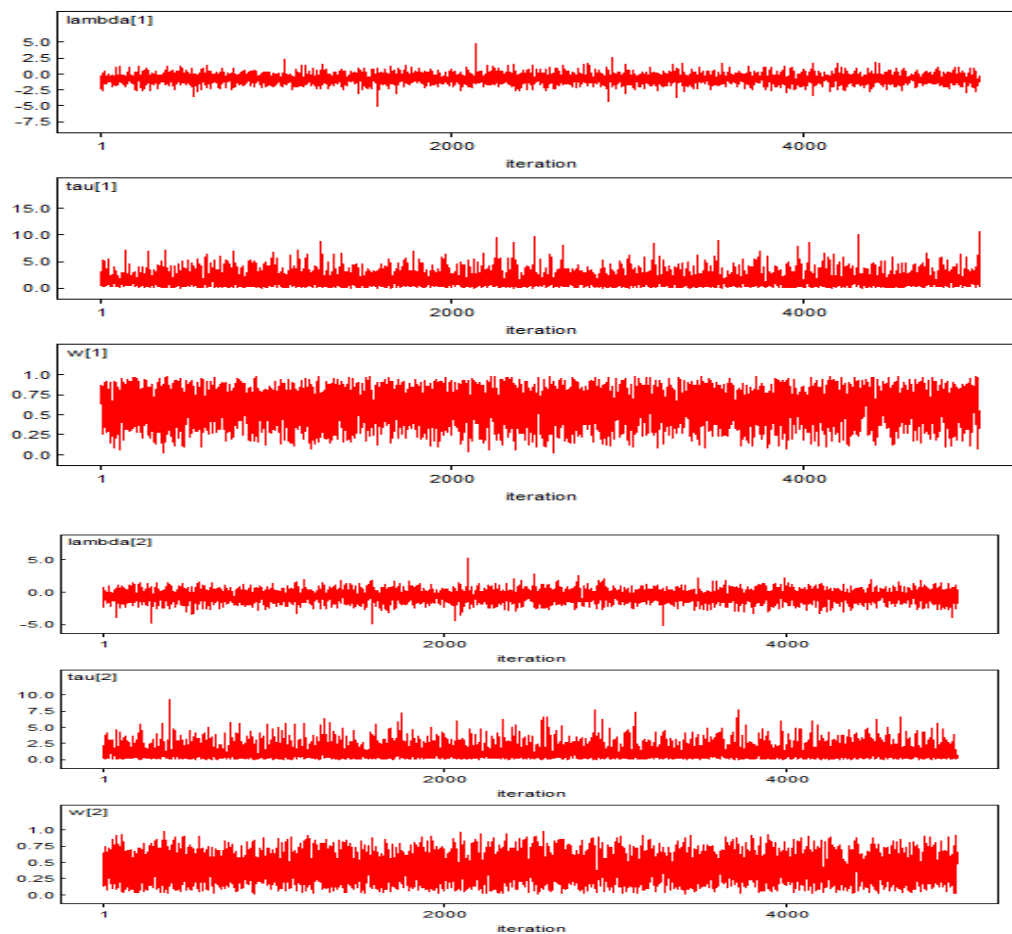


**Gambar L.4.17.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081

## Lampiran 5. (Lanjutan)

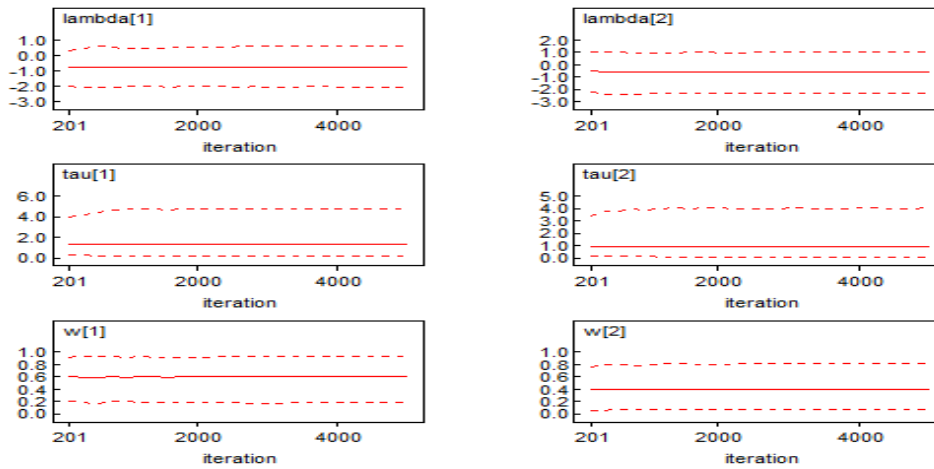


**Gambar L.4.18.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081



**Gambar L.4.19.** History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081

## Lampiran 5. (Lanjutan)

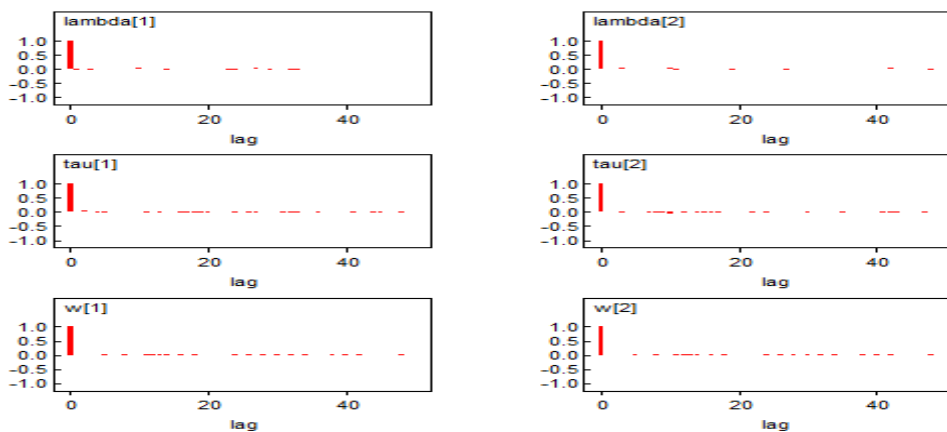


**Gambar L.4.20.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081

### 6. ID Gen LS0752 (Pembangkitan Parameter 5.000 Kali)

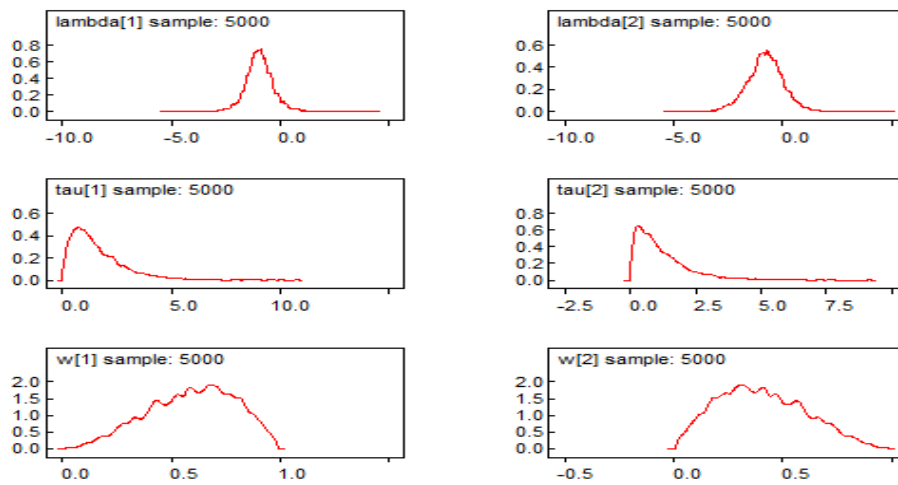
**Tabel L.4.6.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0752

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,9377	0,6418	0,0079	-2,203	-0,9606	0,4015	1	5.000
lambda[2]	-0,8007	0,8281	0,0111	-2,518	-0,7824	0,8116	1	5.000
tau[1]	1,595	1,237	0,0178	0,158	1,278	4,798	1	5.000
tau[2]	1,202	1,041	0,0134	0,0821	0,9101	4,054	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

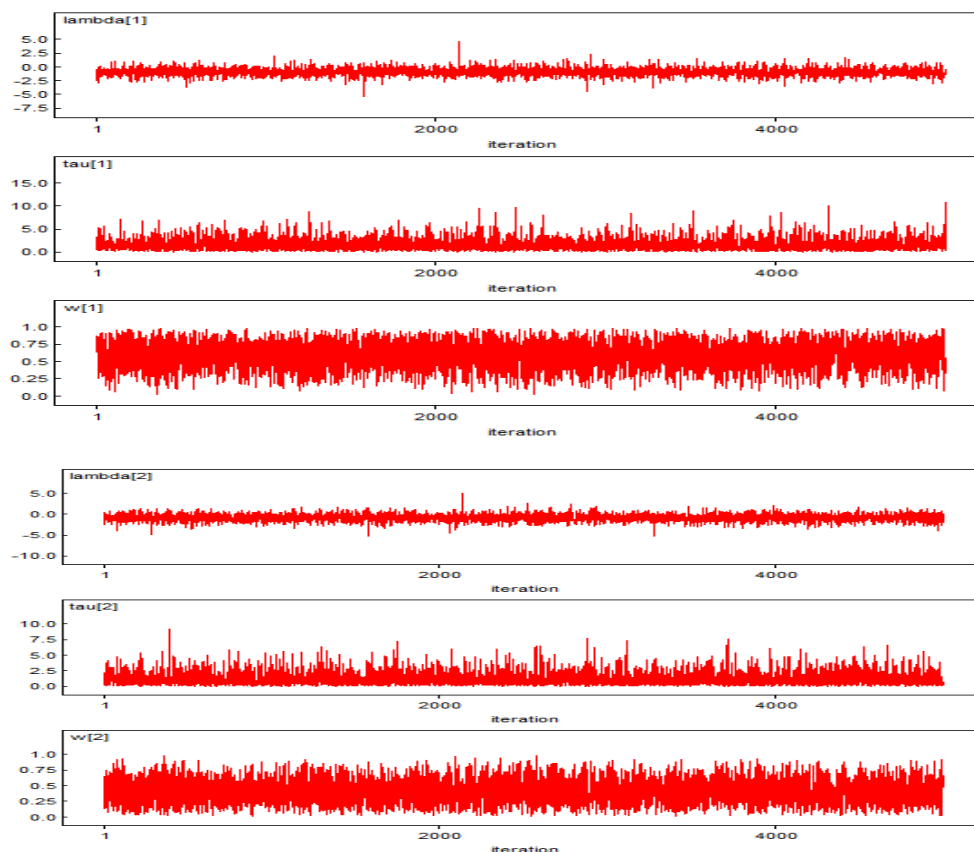


**Gambar L.4.21.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752

## Lampiran 5. (Lanjutan)

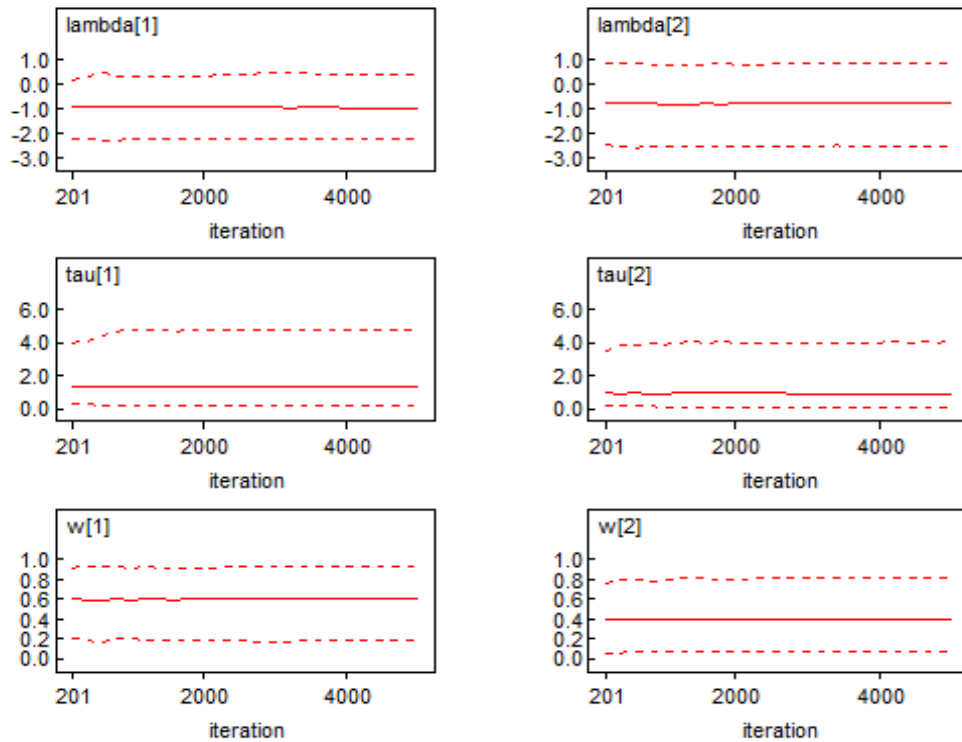


**Gambar L.4.22.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752



**Gambar L.4.23.** History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752

## Lampiran 5. (Lanjutan)



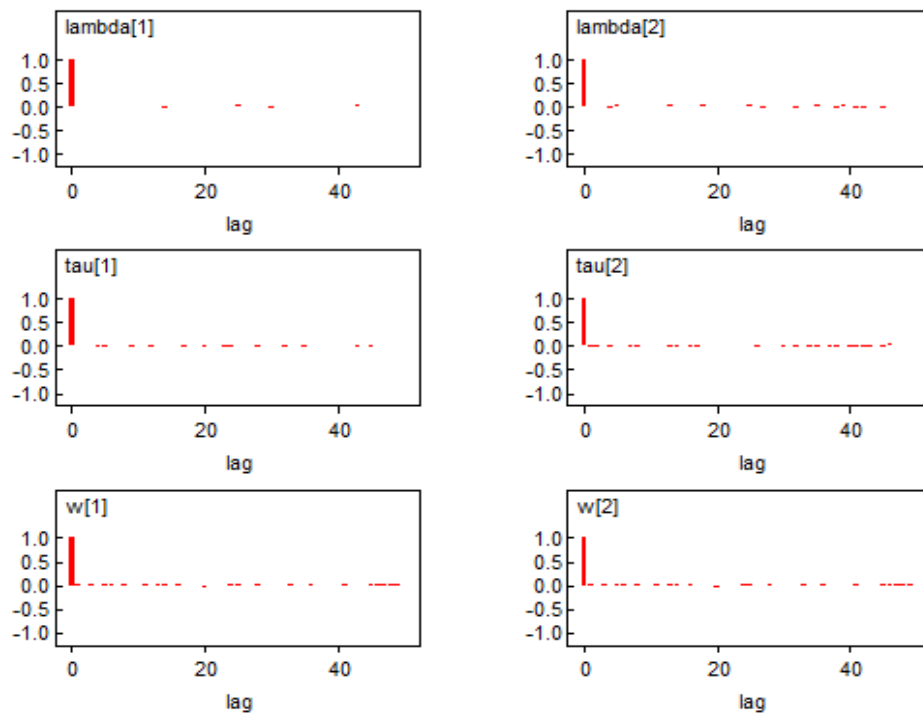
**Gambar L.4.24.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752

## 7. ID Gen LS0688 (Pembangkitan Parameter 5.000 Kali)

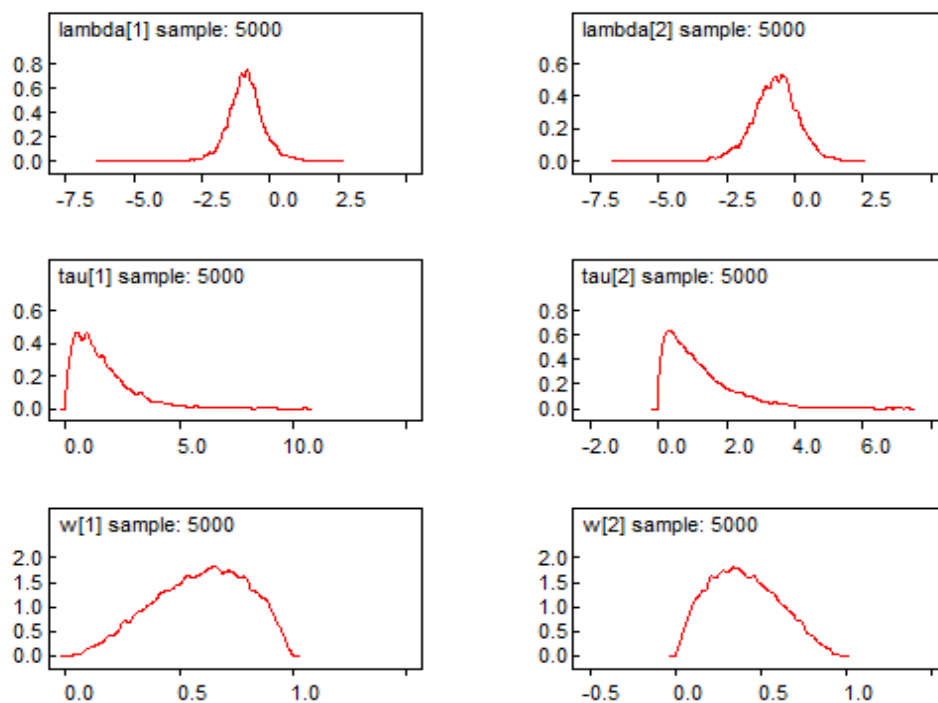
**Tabel L.4.7.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0688

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,8827	0,6643	0,0091	-2,202	-0,8943	0,5438	1	5.000
lambda[2]	-0,7087	0,8325	0,0105	-2,454	-0,6817	0,8784	1	5.000
tau[1]	1,562	1,221	0,0177	0,1468	1,265	4,644	1	5.000
tau[2]	1,194	1,02	0,0144	0,0719	0,9184	3,832	1	5.000
w[1]	0,5957	0,2023	0,0031	0,1879	0,6122	0,9293	1	5.000
w[2]	0,4043	0,2023	0,0031	0,0711	0,3878	0,8122	1	5.000

### Lampiran 5. (Lanjutan)



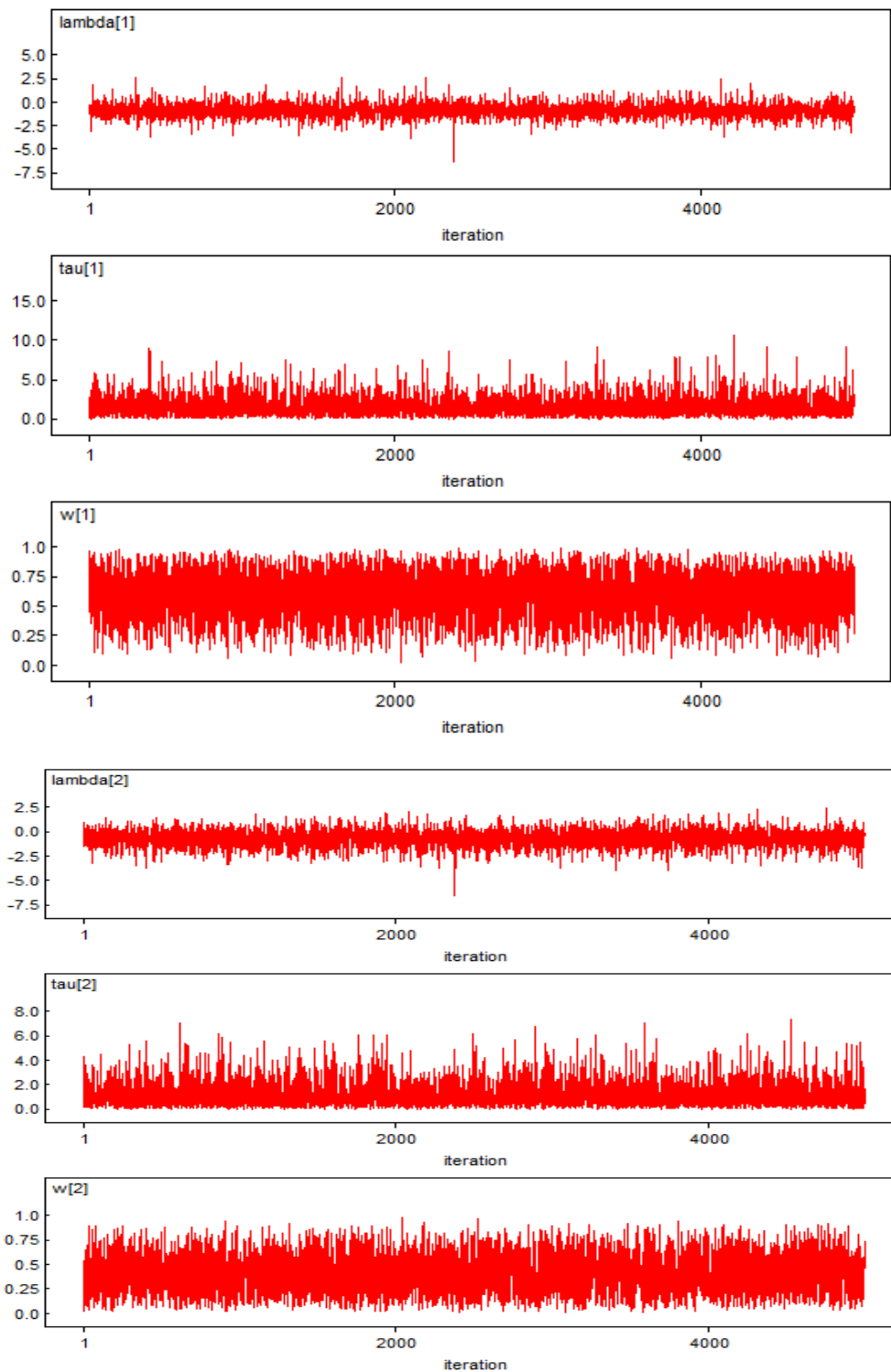
**Gambar L.4.25.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688



**Gambar L.4.26.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688

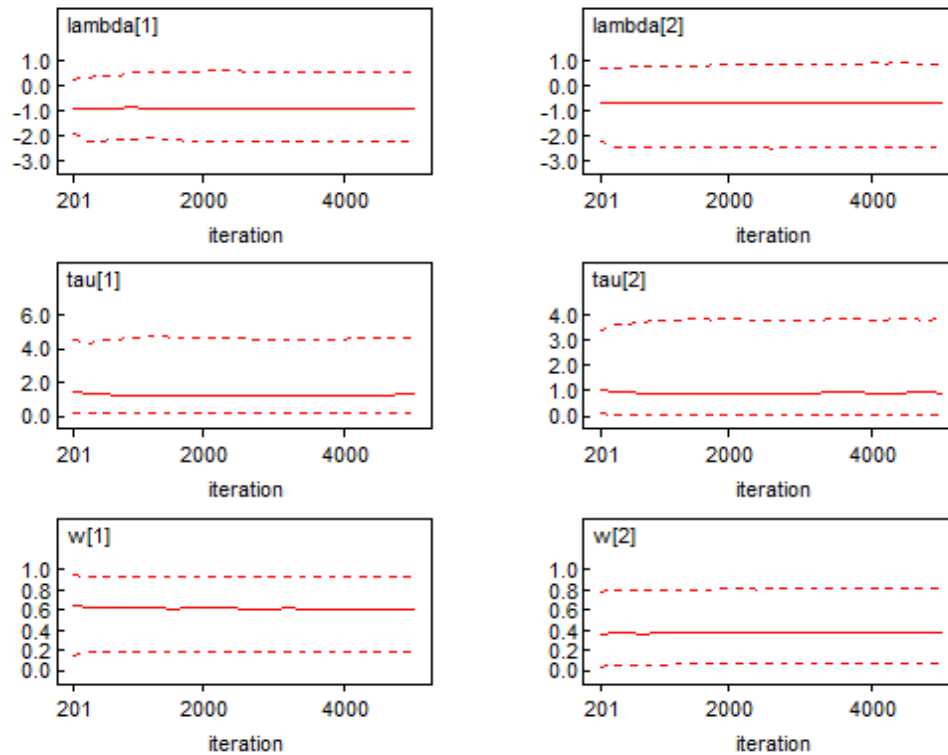


## Lampiran 5. (Lanjutan)



**Gambar L.4.27.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688

## Lampiran 5. (Lanjutan)



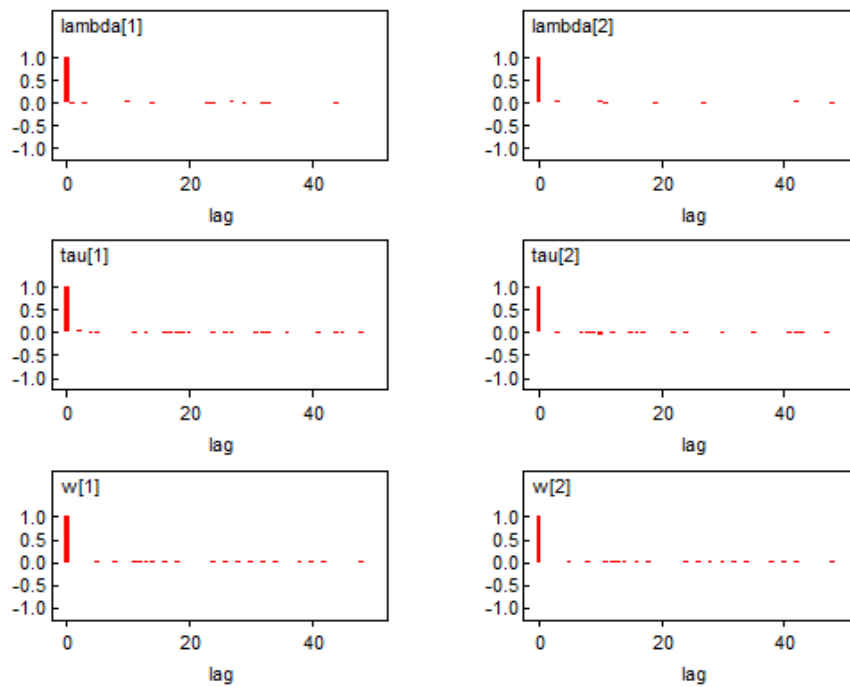
**Gambar L.4.28.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688

## 8. ID Gen U273 (Pembangkitan Parameter 5.000 Kali)

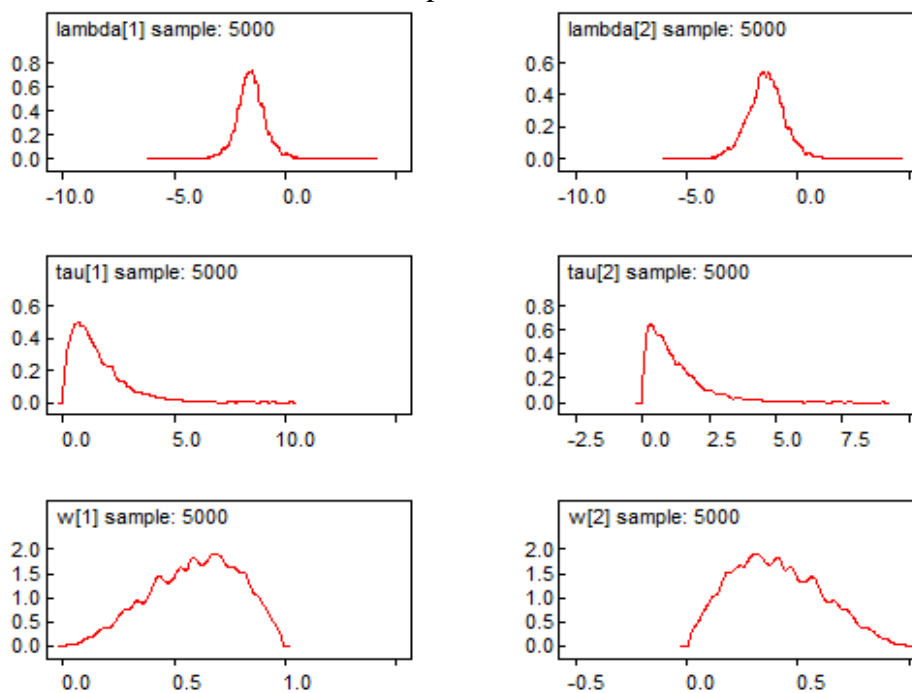
**Tabel L.4.8.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U273

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,526	0,6526	0,0081	-2,82	-1,546	-0,1779	1	5.000
lambda[2]	-1,44	0,8283	0,0111	-3,131	-1,432	0,1925	1	5.000
tau[1]	1,513	1,17	0,0168	0,1494	1,209	4,553	1	5.000
tau[2]	1,206	1,042	0,0135	0,0832	0,9152	4,068	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 5. (Lanjutan)

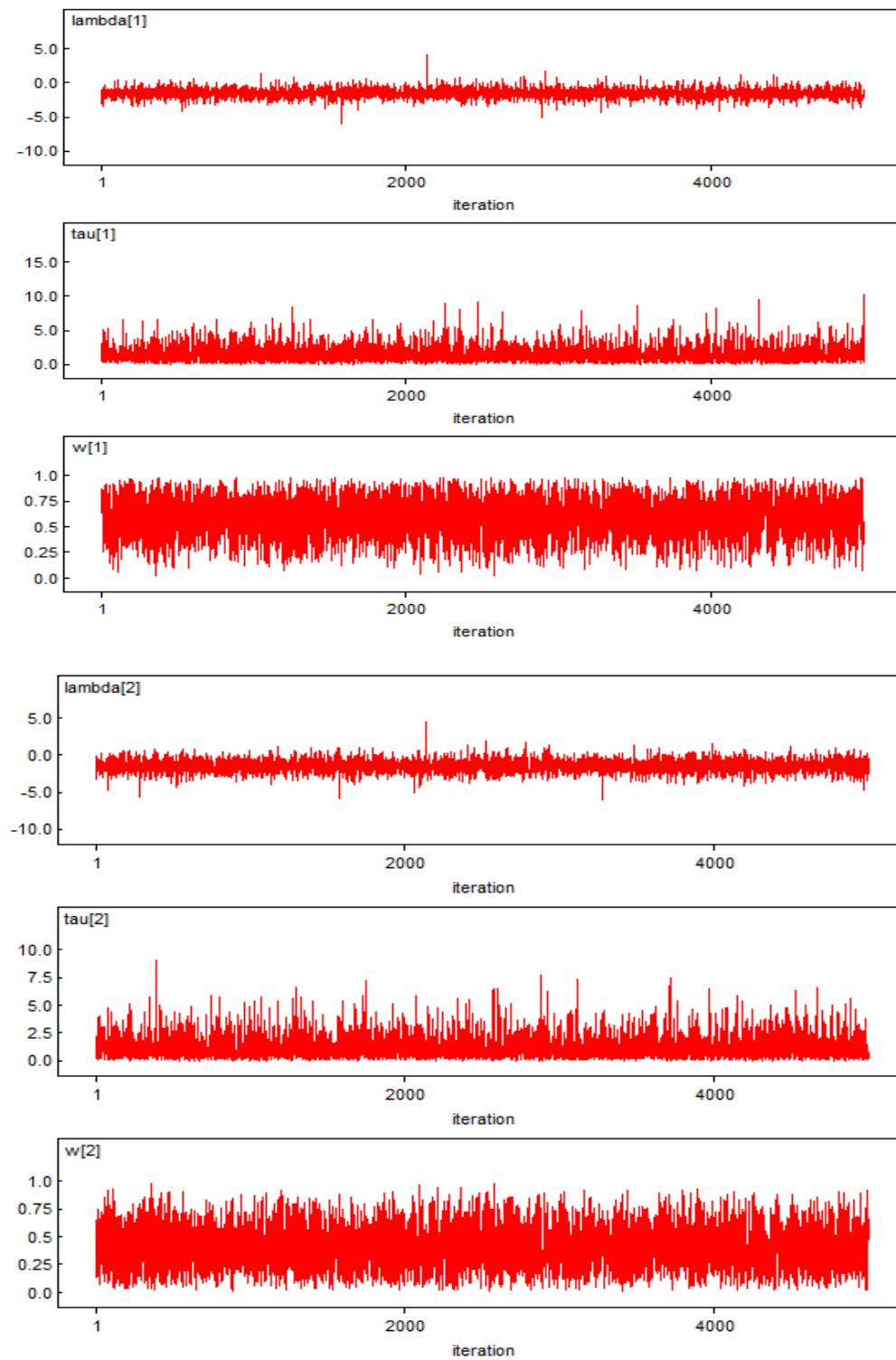


**Gambar L.4.29.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U273



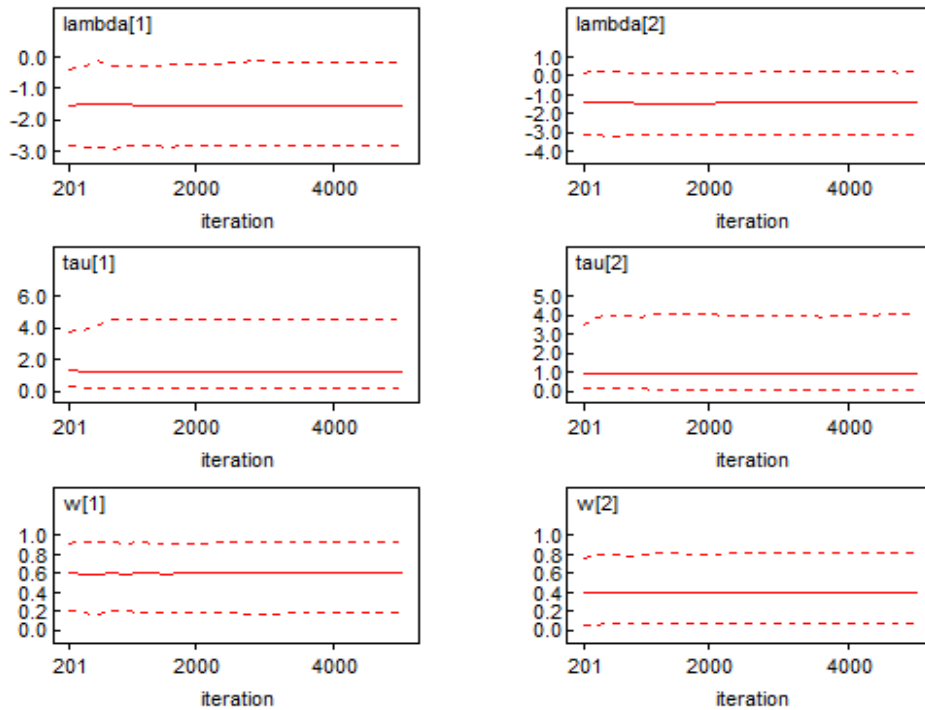
**Gambar L.4.30.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U273

## Lampiran 5. (Lanjutan)



**Gambar L.4.30.** *History Data* Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U273

## Lampiran 5. (Lanjutan)



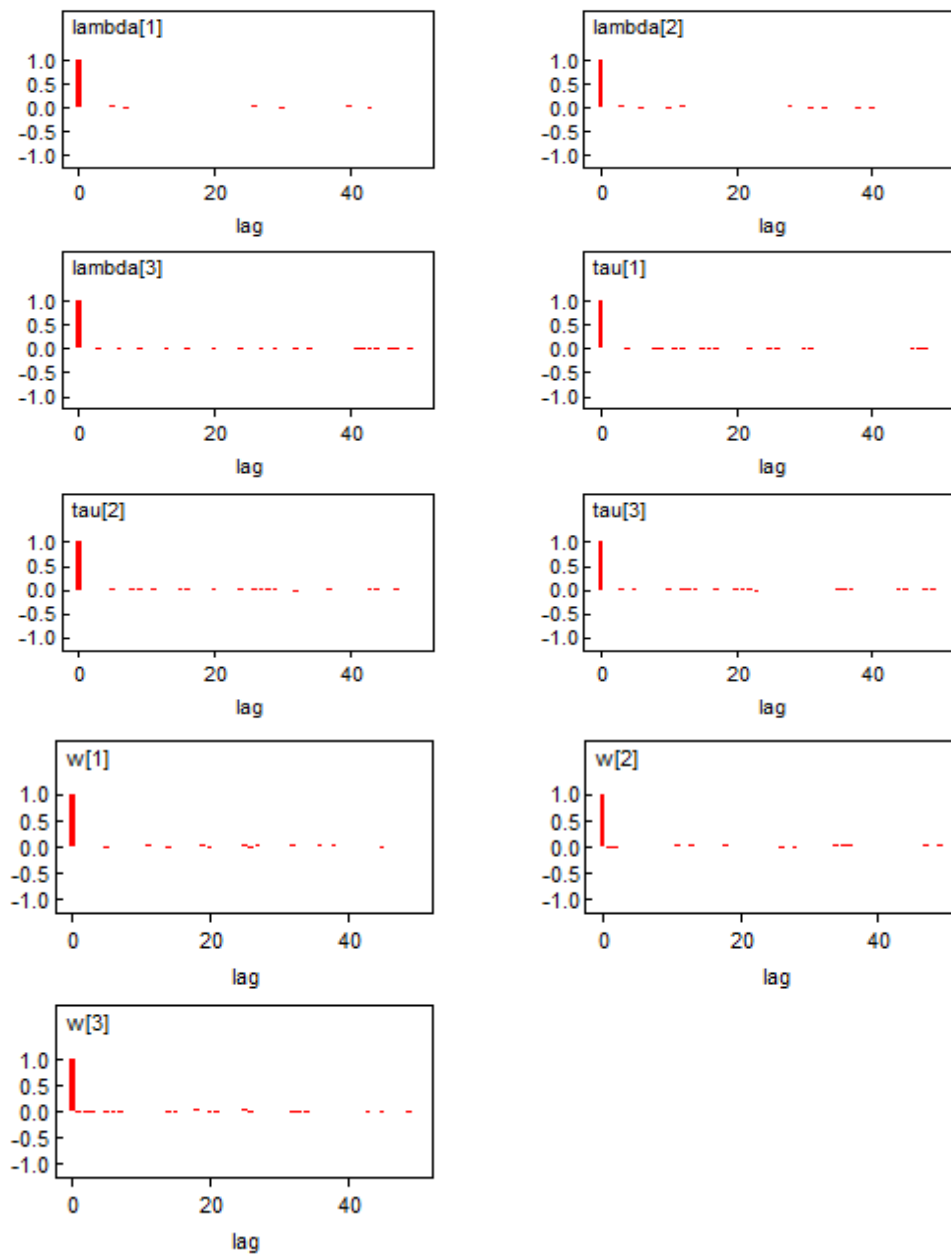
**Gambar L.4.31.** *Quantiles* setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U273

## 9. ID Gen U070 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.4.9.** Estimasi Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070

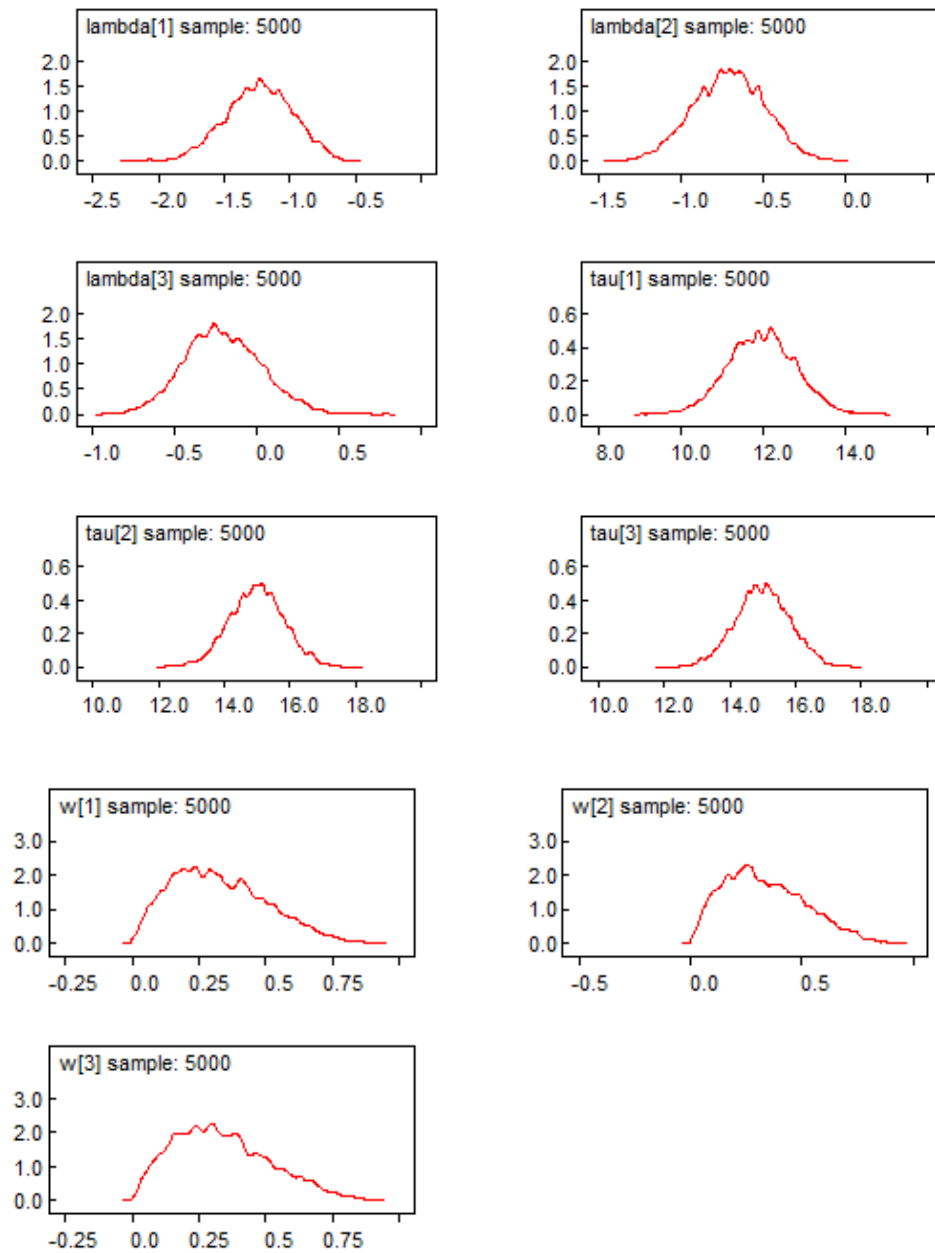
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,233	0,2544	0,0039	-1,743	-1,227	-0,7531	1	5.000
lambda[2]	-0,7039	0,219	0,0031	-1,13	-0,7011	-0,287	1	5.000
lambda[3]	-0,2081	0,235	0,0033	-0,6529	-0,2196	0,2709	1	5.000
tau[1]	11,98	0,8046	0,0107	10,42	11,98	13,55	1	5.000
tau[2]	15	0,8278	0,0107	13,37	15,01	16,65	1	5.000
tau[3]	15,01	0,8286	0,0118	13,38	15,01	16,62	1	5.000
w[1]	0,3276	0,1754	0,0023	0,0515	0,3067	0,7026	1	5.000
w[2]	0,3348	0,1801	0,0024	0,0556	0,3121	0,723	1	5.000
w[3]	0,3376	0,1787	0,0023	0,0544	0,3162	0,7217	1	5.000

### Lampiran 5. (Lanjutan)



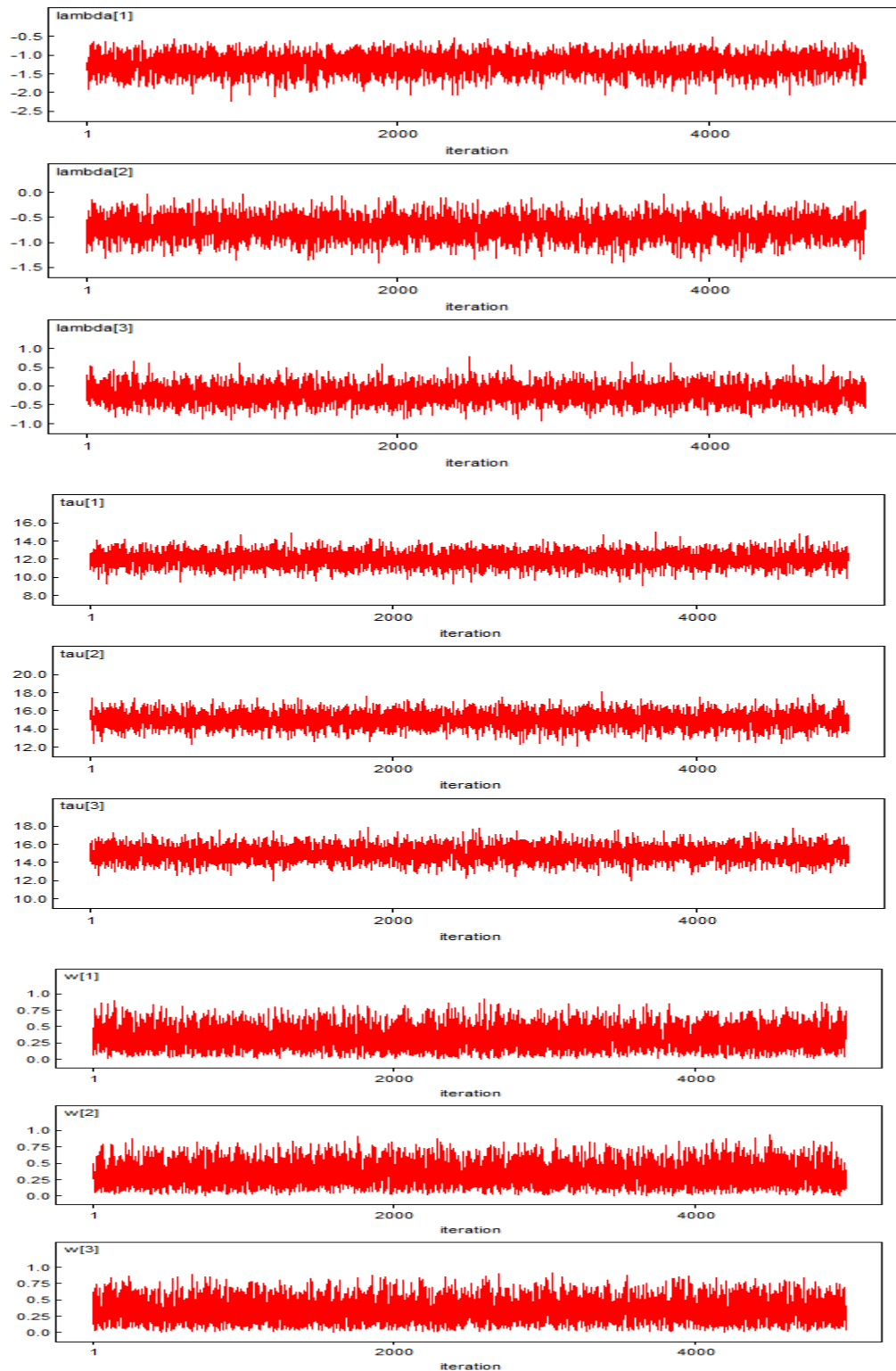
**Gambar L.4.32.** Fungsi Autokorelasi setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070

## Lampiran 5. (Lanjutan)



**Gambar L.4.33.** Densitas Kernel setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070

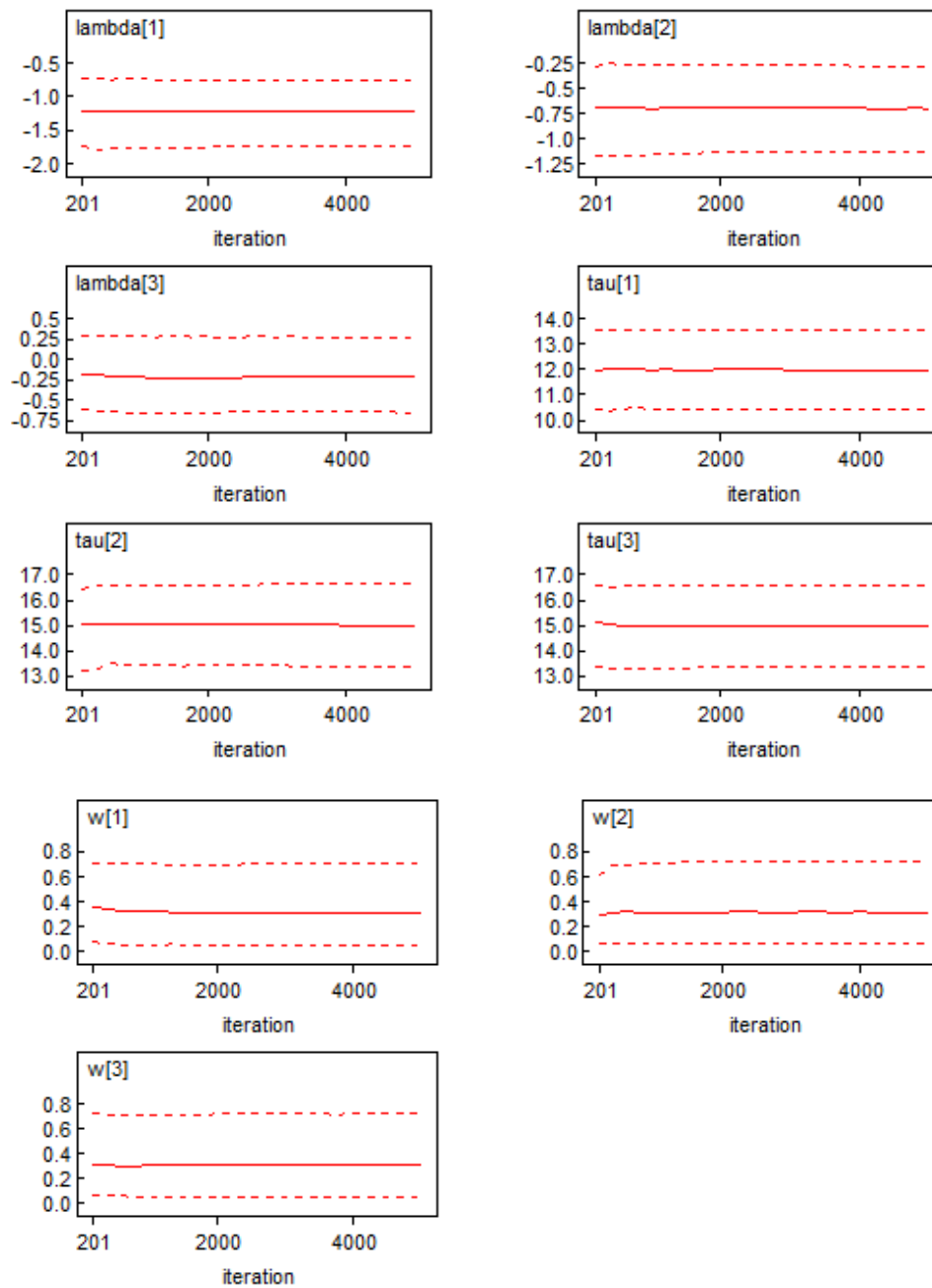
## Lampiran 5. (Lanjutan)



**Gambar L.4.34.** *History Data* Bangkitan setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070



## Lampiran 5. (Lanjutan)



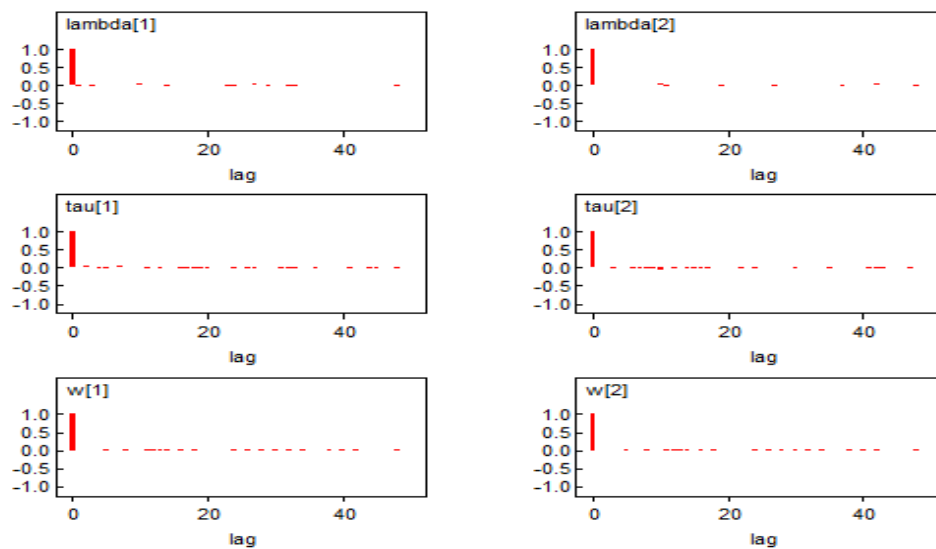
**Gambar L.4.35.** *Quantiles* setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070

## Lampiran 5. (Lanjutan)

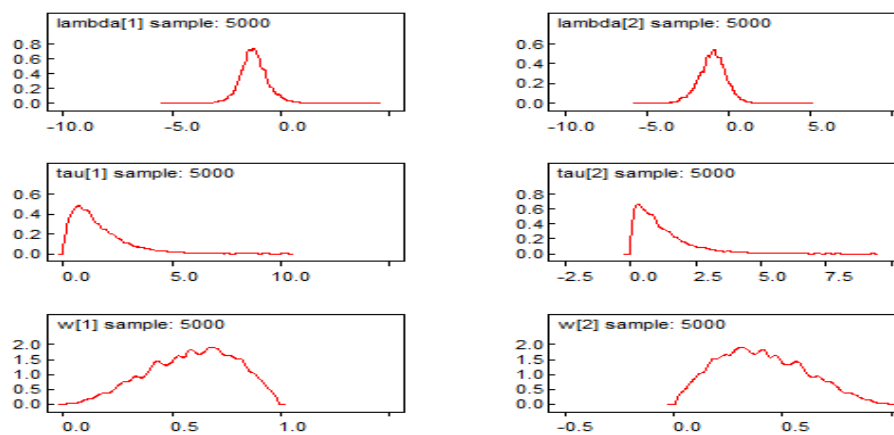
### 10. ID Gen U325 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.4.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U325

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,223	0,655	0,0079	-2,448	-1,253	0,1979	1	5.000
lambda[2]	-0,9717	0,8445	0,0112	-2,753	-0,9429	0,6421	1	5.000
tau[1]	1,573	1,229	0,0179	0,1501	1,254	4,774	1	5.000
tau[2]	1,181	1,033	0,0134	0,081	0,8886	4,005	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

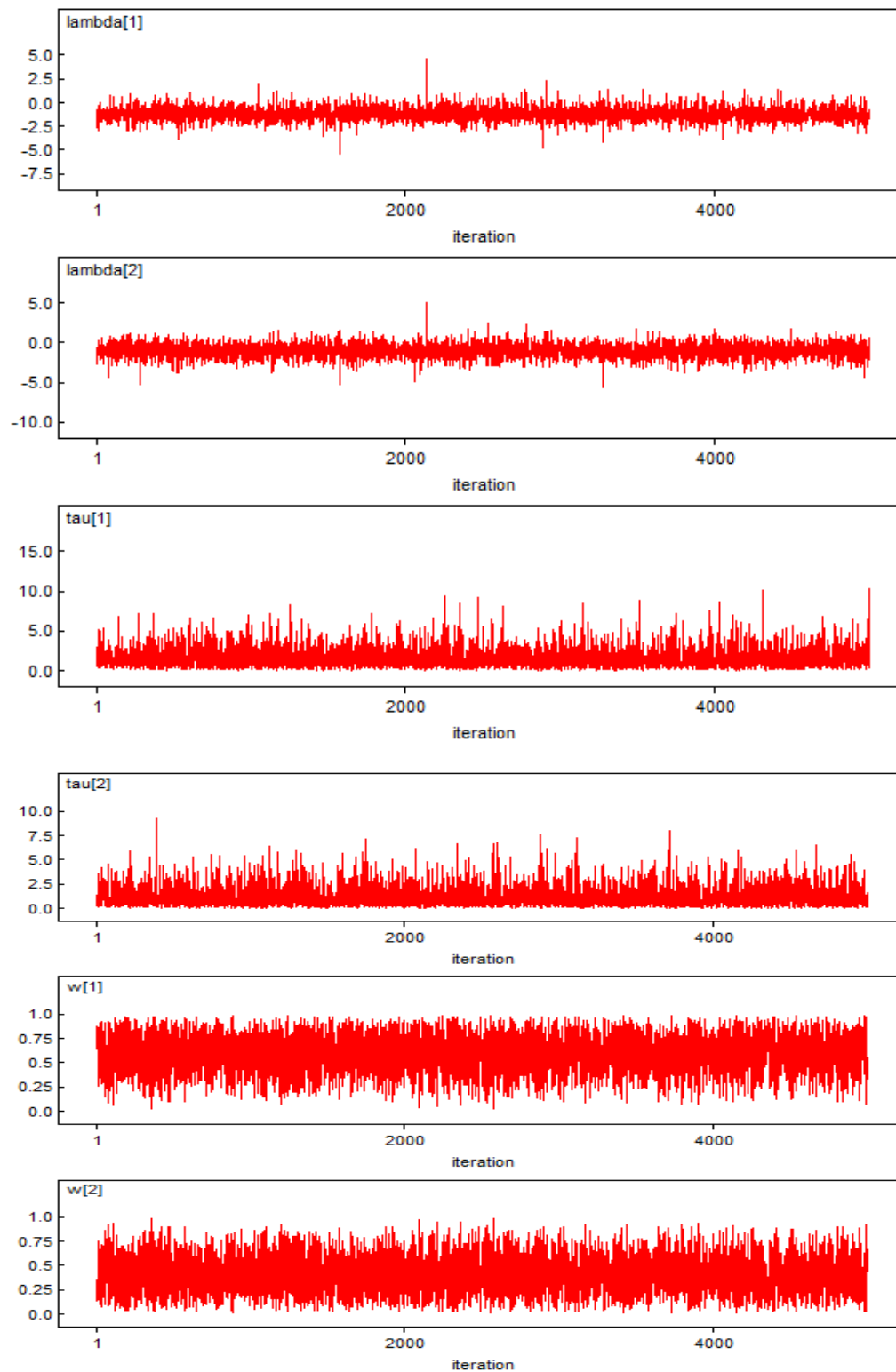


**Gambar L.4.36.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325



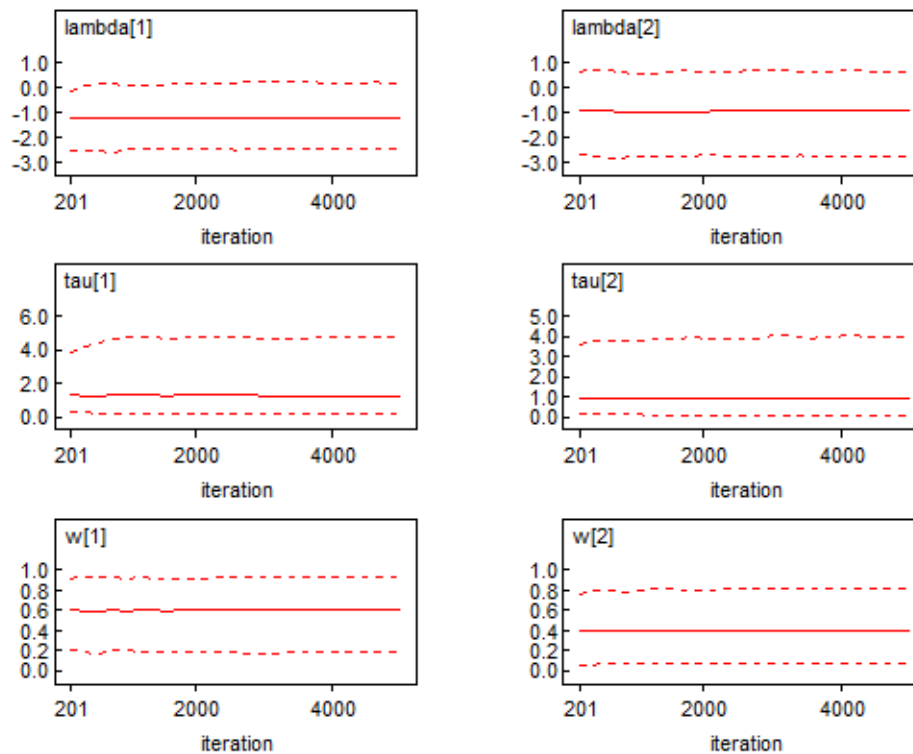
**Gambar L.4.37.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325

## Lampiran 5. (Lanjutan)



**Gambar L.4.38.** *History Data Bangkitan* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325

## Lampiran 5. (Lanjutan)



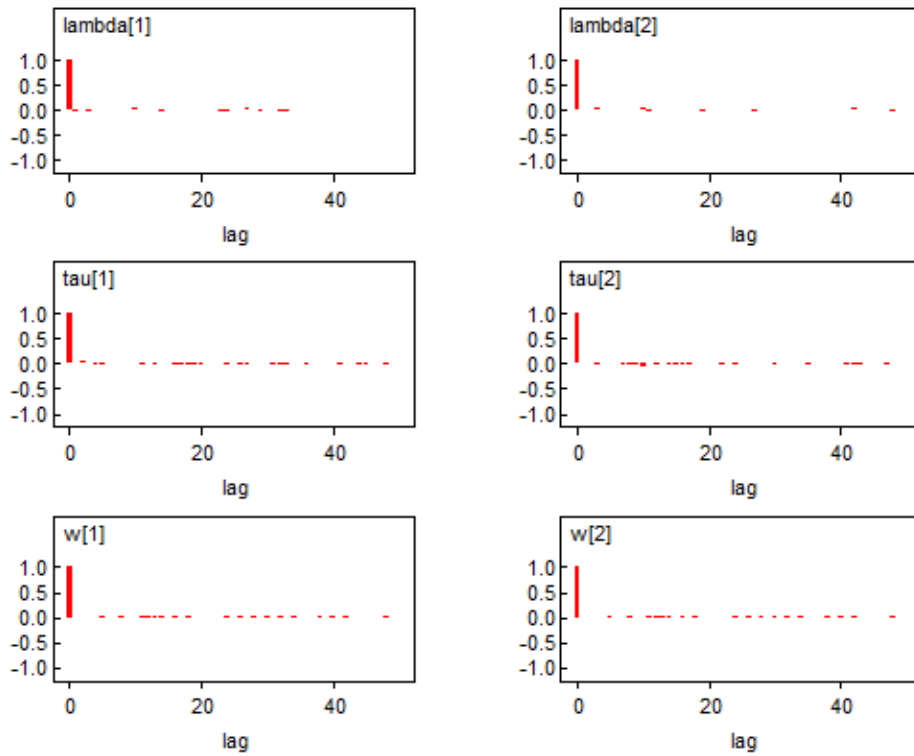
**Gambar L.4.39.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325

### 11. ID Gen U385 (Pembangkitan Parameter 5.000 Kali)

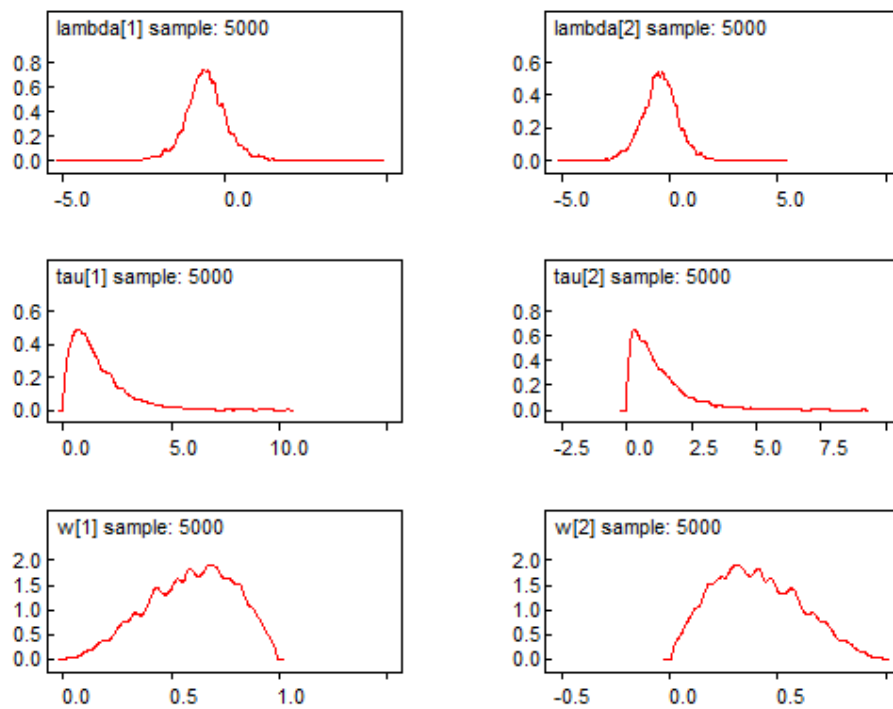
**Tabel L.4.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U385

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5845	0,6488	0,008	-1,872	-0,6047	0,7496	1	5.000
lambda[2]	-0,4571	0,8295	0,0111	-2,171	-0,4397	1,167	1	5.000
tau[1]	1,535	1,188	0,0171	0,1512	1,228	4,622	1	5.000
tau[2]	1,203	1,041	0,0134	0,0821	0,91	4,056	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

### Lampiran 5. (Lanjutan)

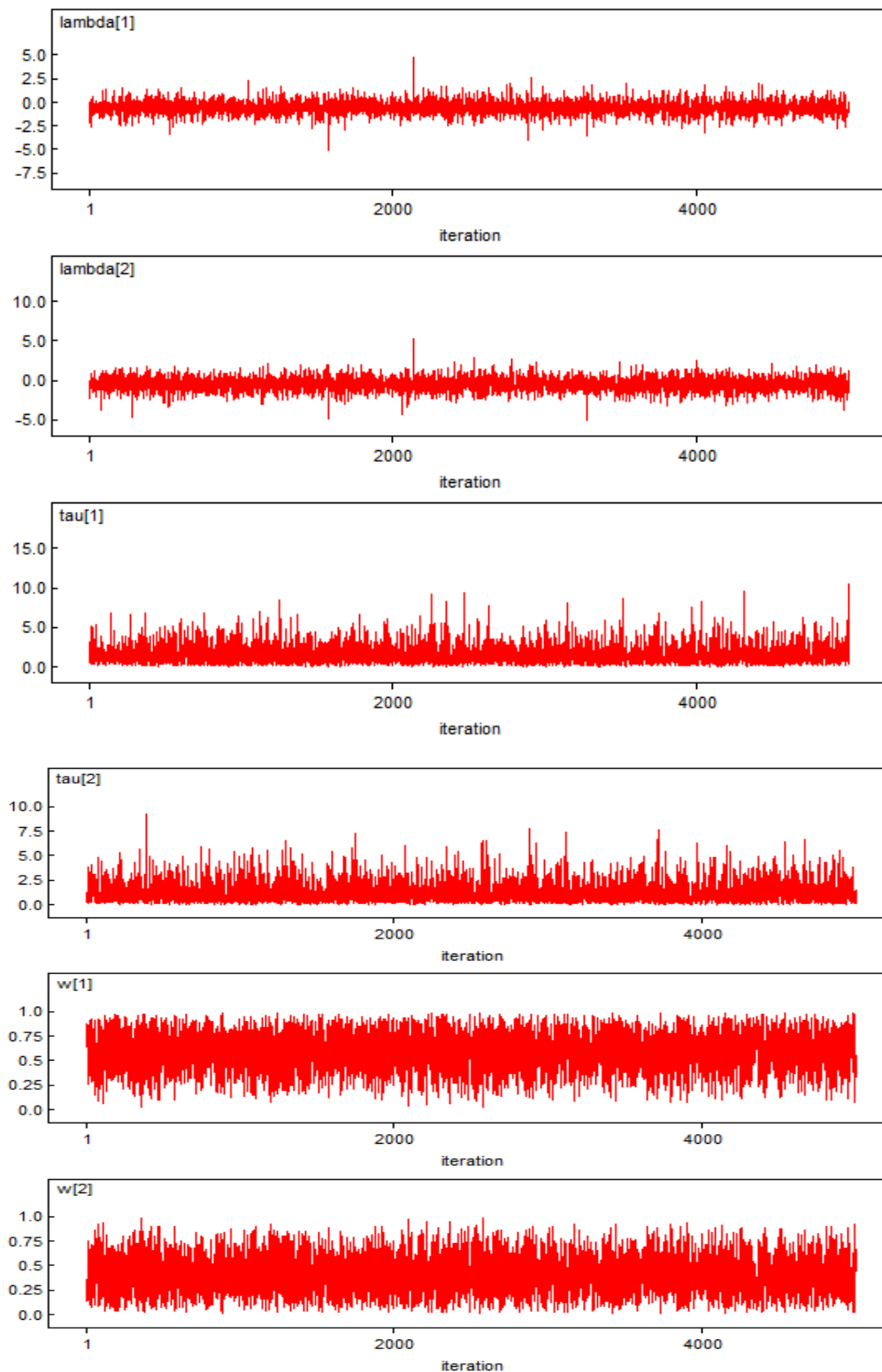


**Gambar L.4.40.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385



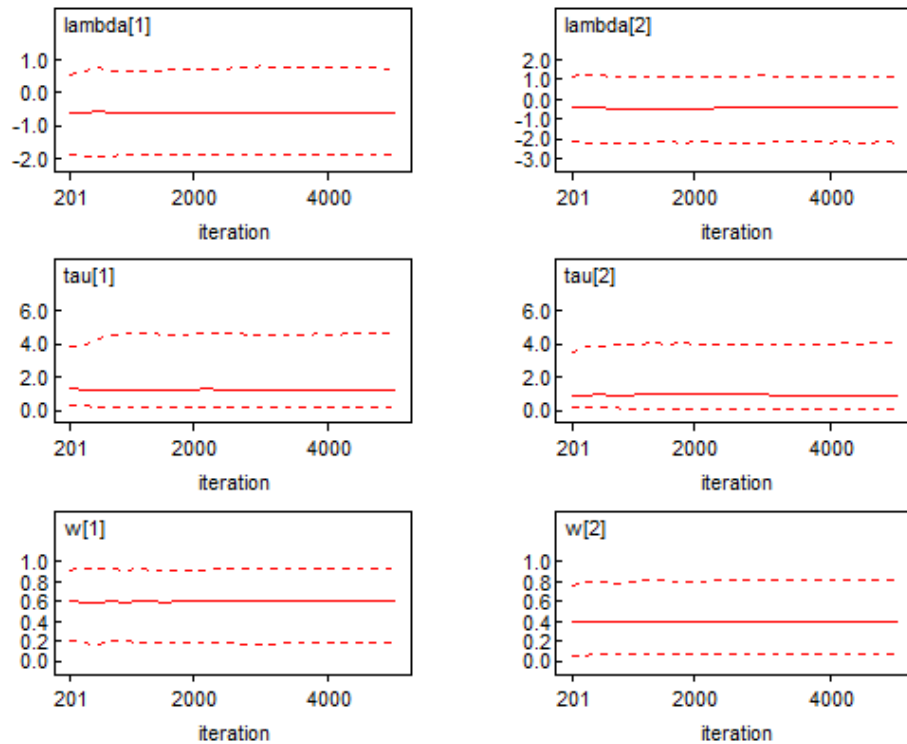
**Gambar L.4.41.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385

## Lampiran 5. (Lanjutan)



**Gambar L.4.42.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385*

## Lampiran 5. (Lanjutan)



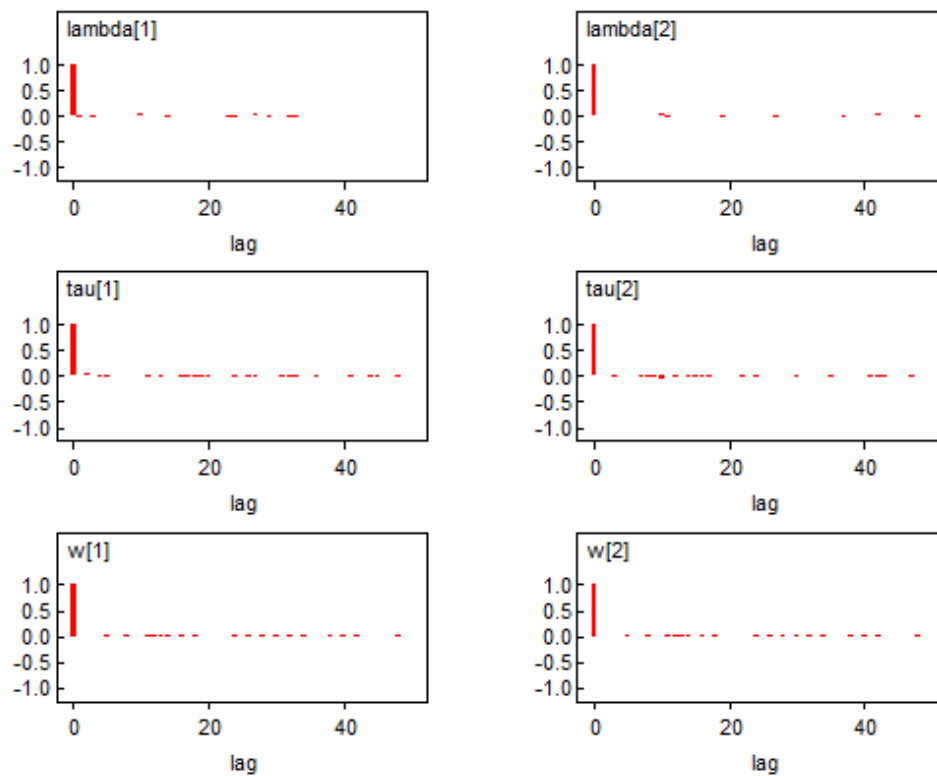
**Gambar L.4.43.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385

## 12. ID Gen U447 (Pembangkitan Parameter 5.000 Kali)

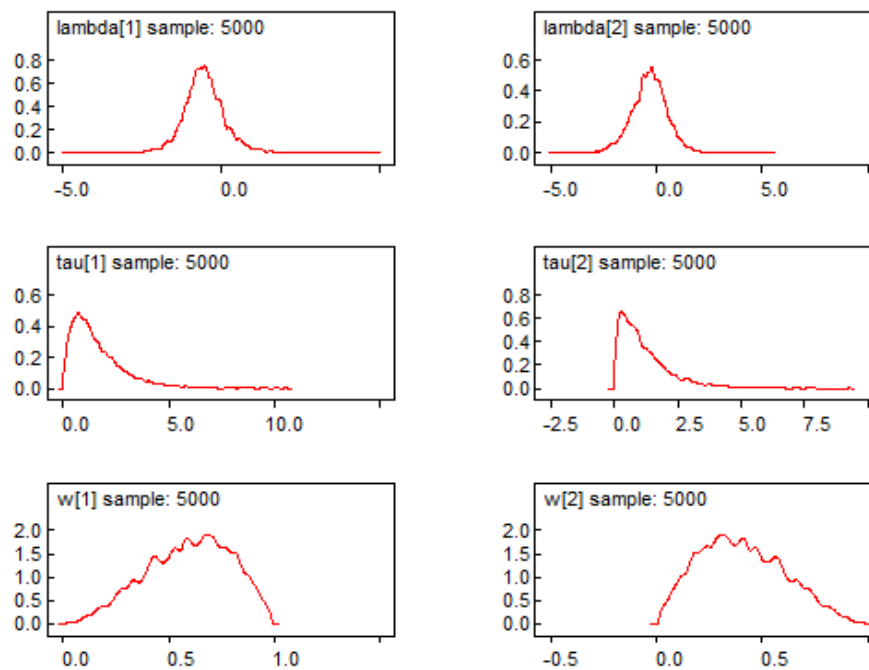
**Tabel L.4.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U447

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
$\lambda[1]$	-0,547	0,6464	0,0079	-1,802	-0,5691	0,8311	1	5.000
$\lambda[2]$	-0,3395	0,8362	0,0111	-2,092	-0,3142	1,273	1	5.000
$\tau[1]$	1,587	1,234	0,0179	0,1563	1,269	4,76	1	5.000
$\tau[2]$	1,191	1,037	0,0134	0,082	0,8995	4,003	1	5.000
$w[1]$	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
$w[2]$	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

### Lampiran 5. (Lanjutan)



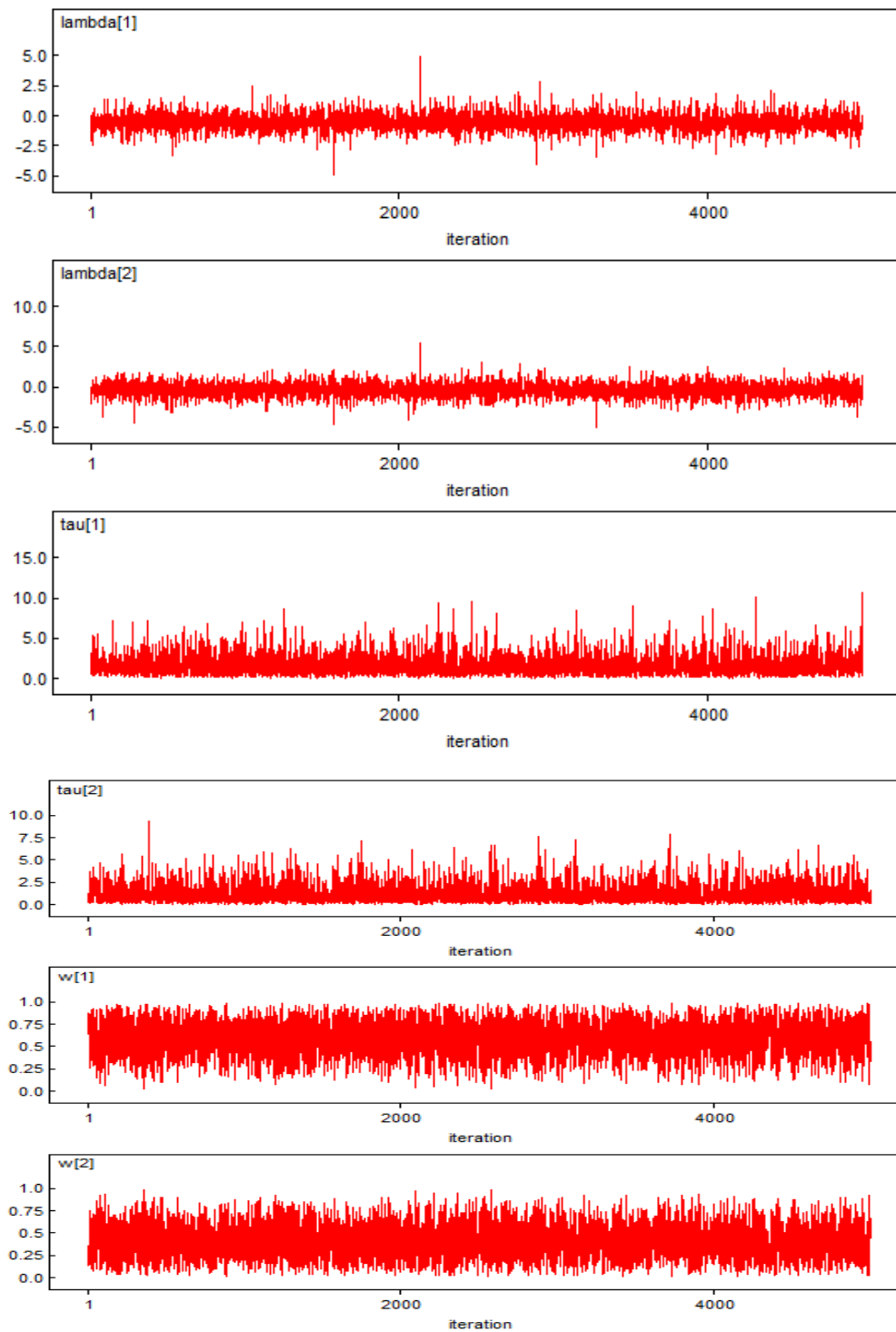
**Gambar L.4.44.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447



**Gambar L.4.45.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447

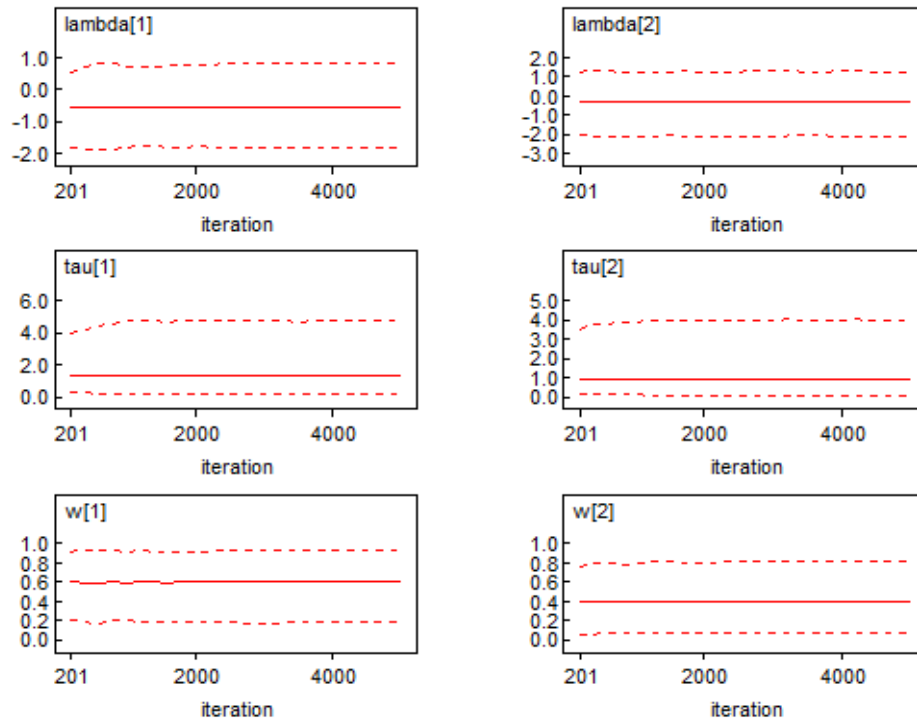


## Lampiran 5. (Lanjutan)



**Gambar L.4.46.** *History Data* Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447

## Lampiran 5. (Lanjutan)



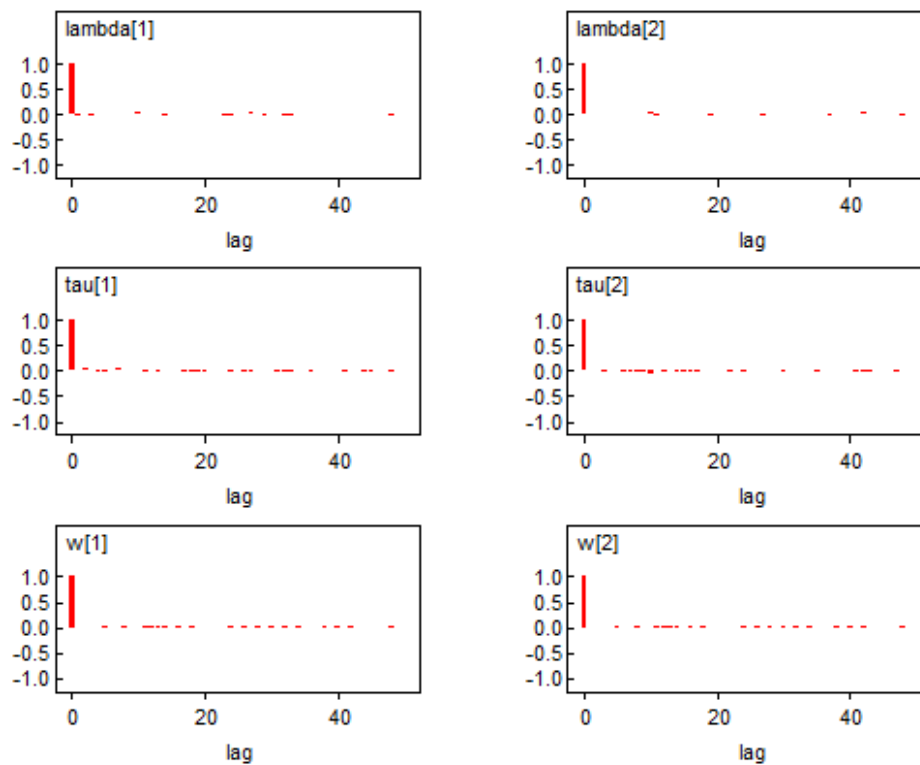
**Gambar L.4.47.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447

### 13. ID Gen CA0426 (Pembangkitan Parameter 5.000 Kali)

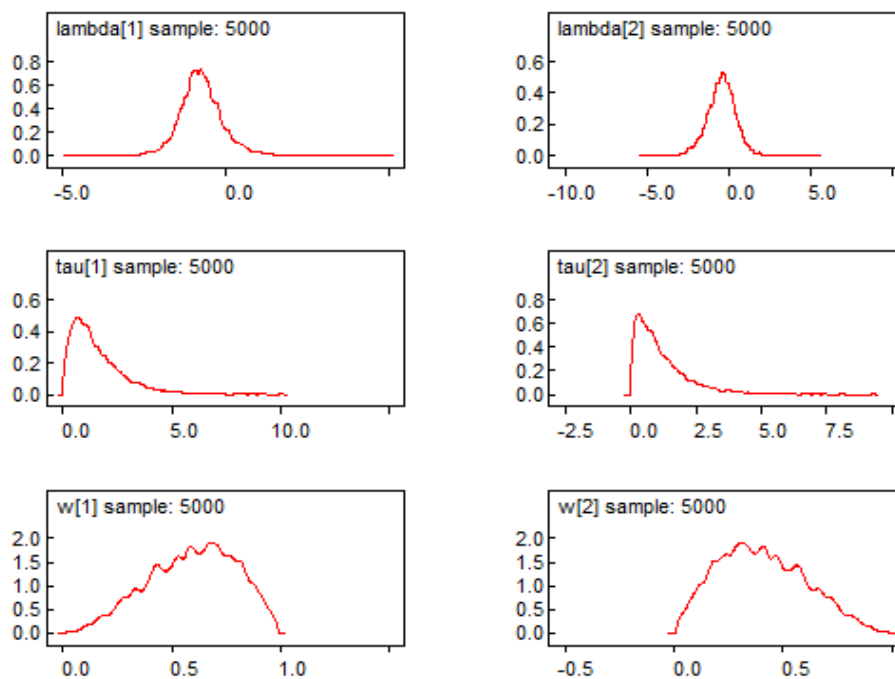
**Tabel L.4.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA0426

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,754	0,6629	0,008	-1,983	-0,788	0,6868	1	5.000
lambda[2]	-0,4422	0,8567	0,0113	-2,267	-0,4035	1,177	1	5.000
tau[1]	1,552	1,216	0,0178	0,1467	1,234	4,731	1	5.000
tau[2]	1,164	1,027	0,0133	0,0802	0,871	3,967	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 5. (Lanjutan)

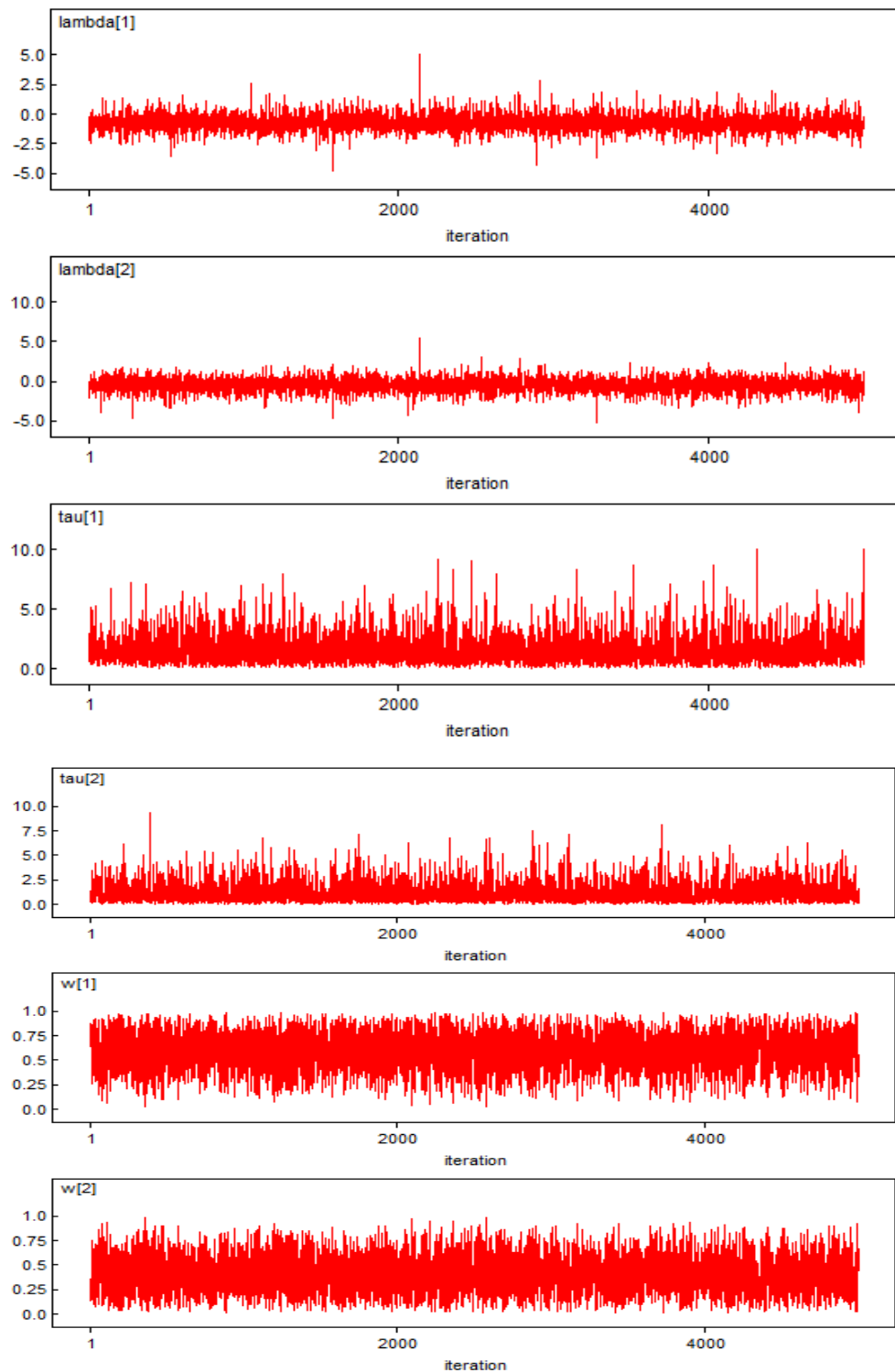


**Gambar L.4.48.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426



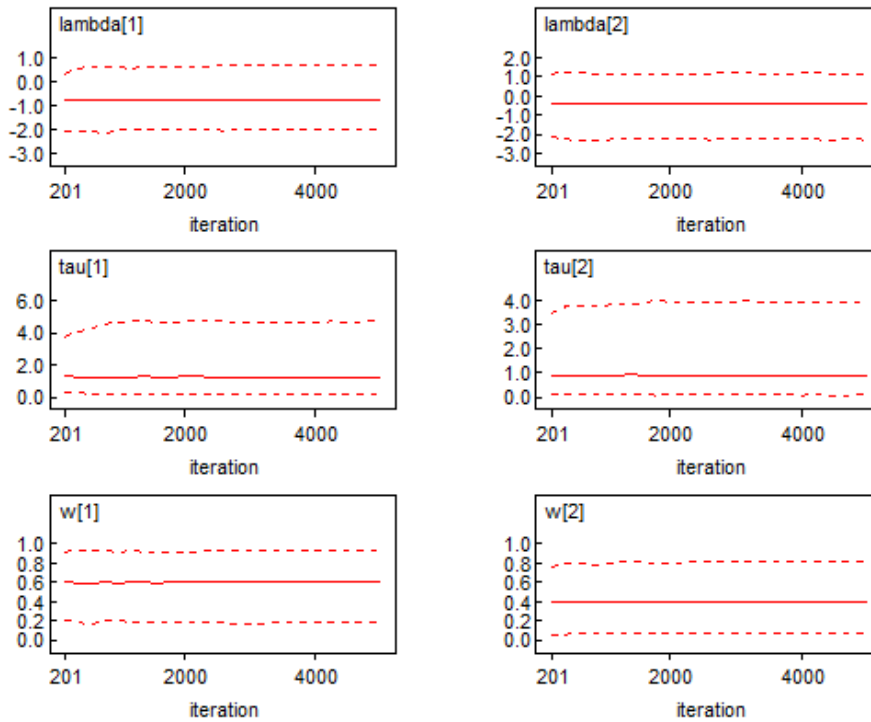
**Gambar L.4.49.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426

### Lampiran 5. (Lanjutan)



**Gambar L.4.50.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426

## Lampiran 5. (Lanjutan)



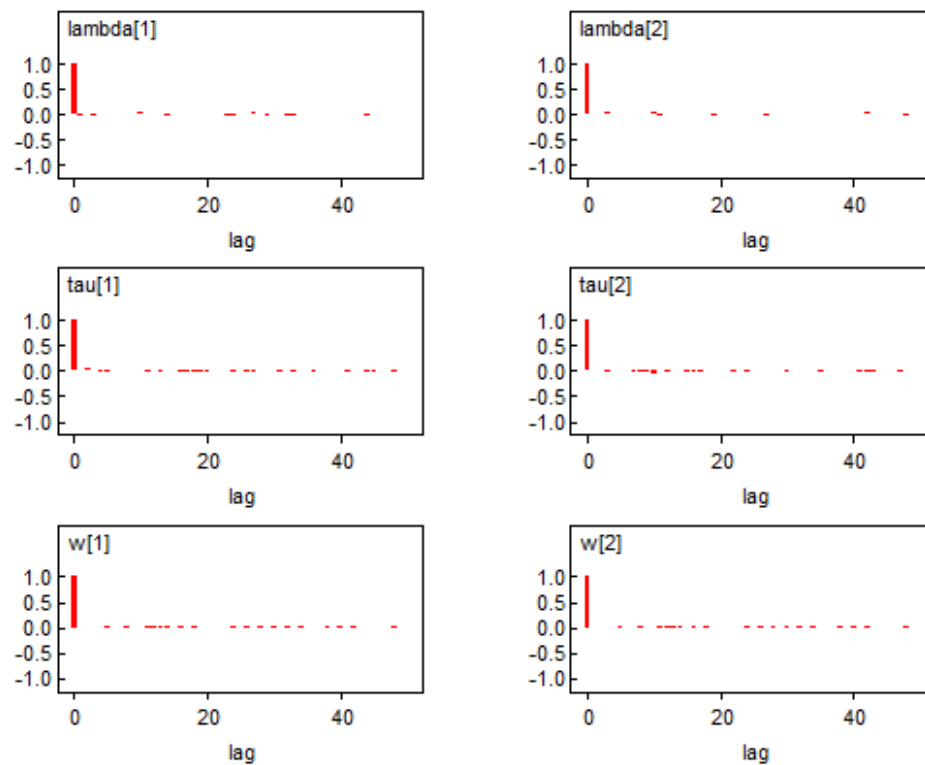
**Gambar L.4.51.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426

### 14. ID Gen CA0534 (Pembangkitan Parameter 5.000 Kali)

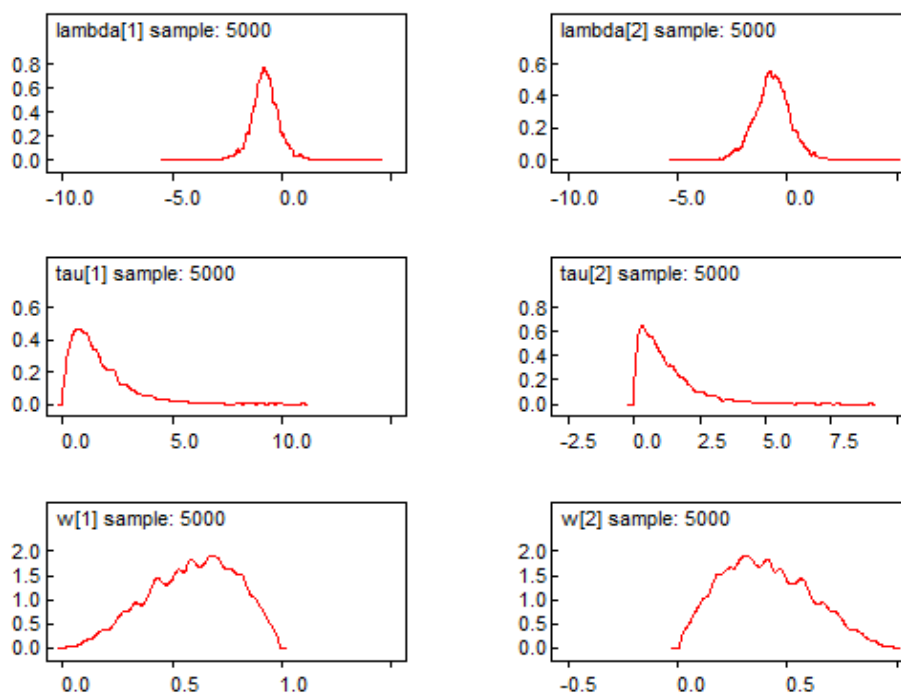
**Tabel L.4.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA0534

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,7472	0,6378	0,0079	-2,032	-0,7602	0,5501	1	5.000
lambda[2]	-0,6854	0,8227	0,0111	-2,359	-0,6813	0,9462	1	5.000
tau[1]	1,6	1,238	0,0177	0,1576	1,28	4,803	1	5.000
tau[2]	1,209	1,043	0,0135	0,0833	0,9192	4,051	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

### Lampiran 5. (Lanjutan)

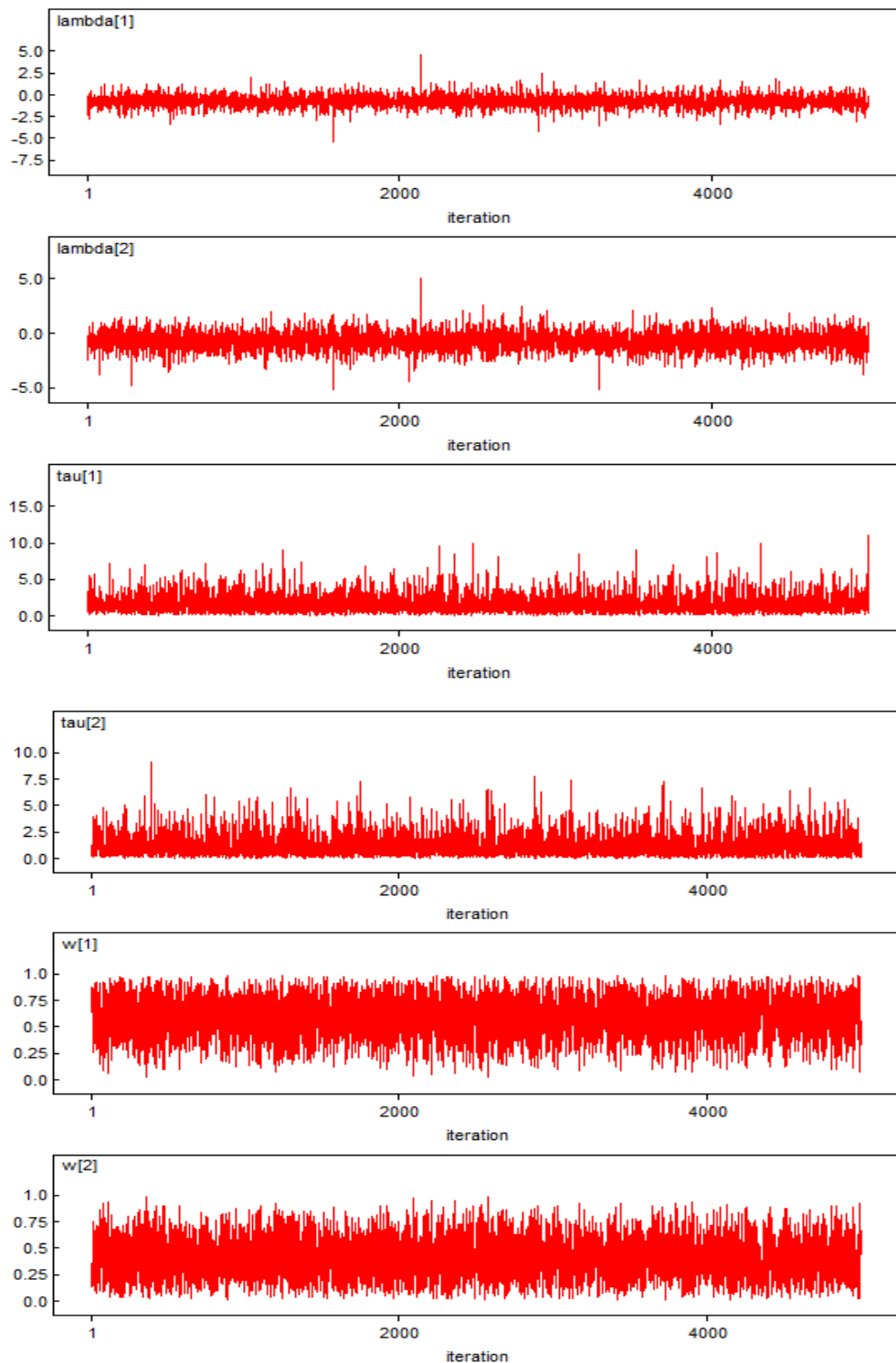


**Gambar L.4.52.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534



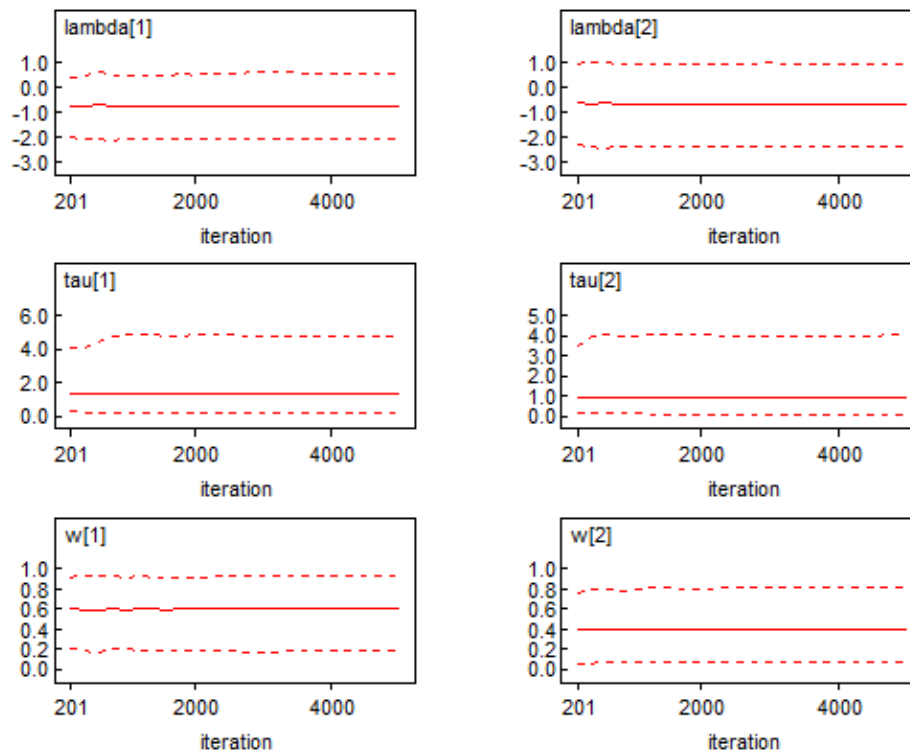
**Gambar L.4.53.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534

## Lampiran 5. (Lanjutan)



**Gambar L.4.54.** *History Data Bangkitan* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534

## Lampiran 5. (Lanjutan)



**Gambar L.4.55.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534

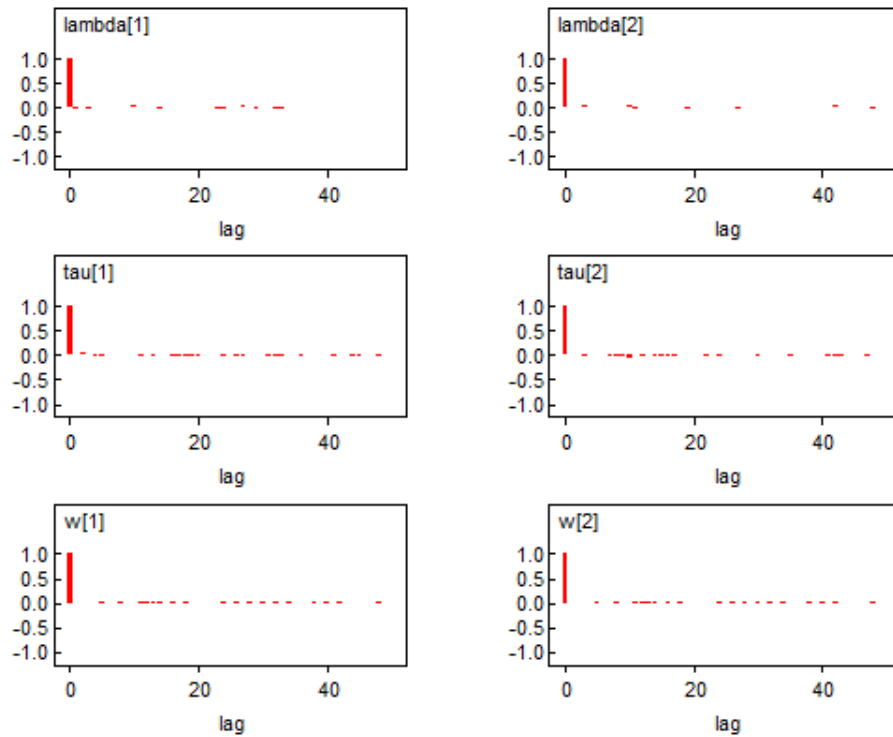
### 15. ID Gen CA0702 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.4.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA0702

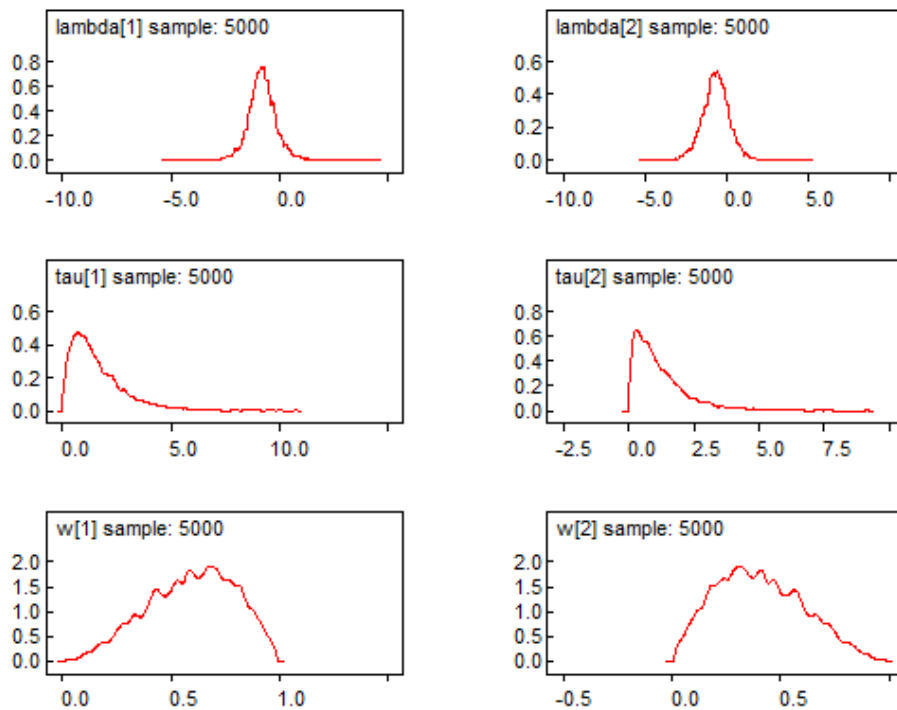
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,7869	0,6416	0,0079	-2,051	-0,8086	0,5535	1	5.000
lambda[2]	-0,6551	0,8277	0,0111	-2,368	-0,6378	0,9573	1	5.000
tau[1]	1,592	1,234	0,0178	0,1573	1,273	4,797	1	5.000
tau[2]	1,203	1,041	0,0134	0,0819	0,9107	4,051	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000



### Lampiran 5. (Lanjutan)

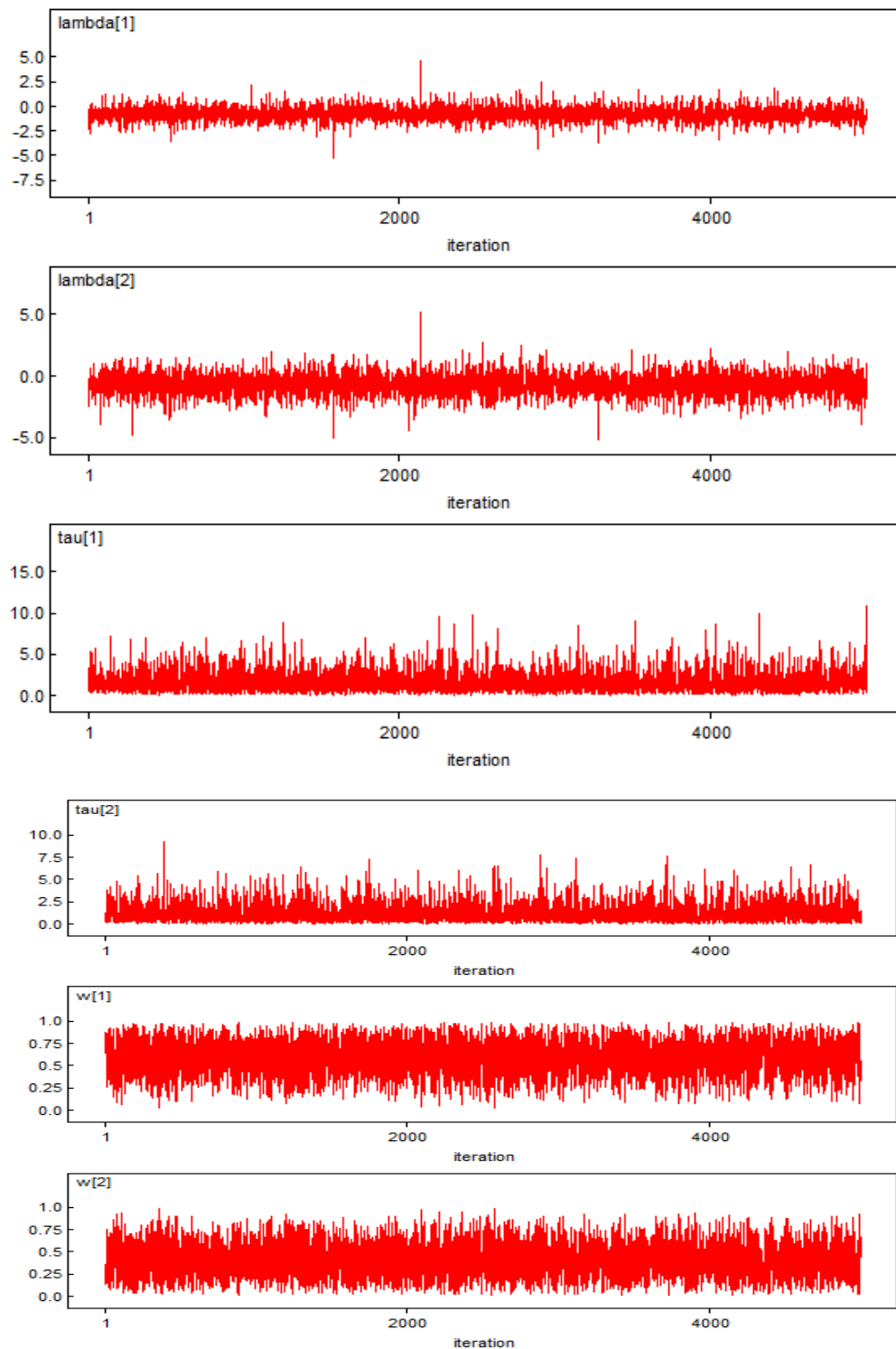


**Gambar L.4.56.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702



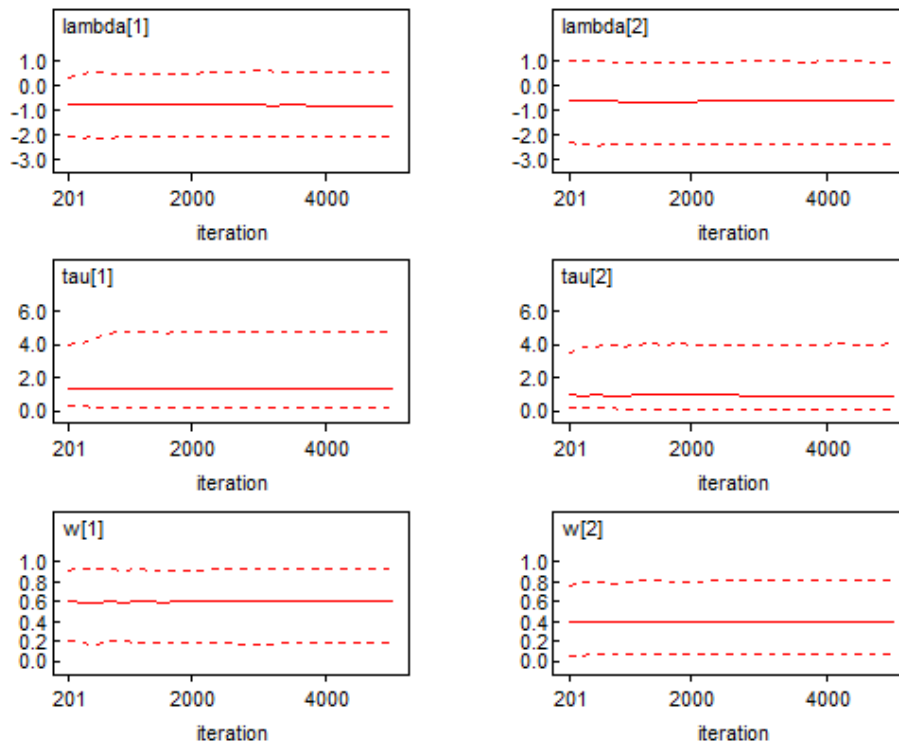
**Gambar L.4.57.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702

## Lampiran 5. (Lanjutan)



**Gambar L.4.58.** *History Data* Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702

## Lampiran 5. (Lanjutan)



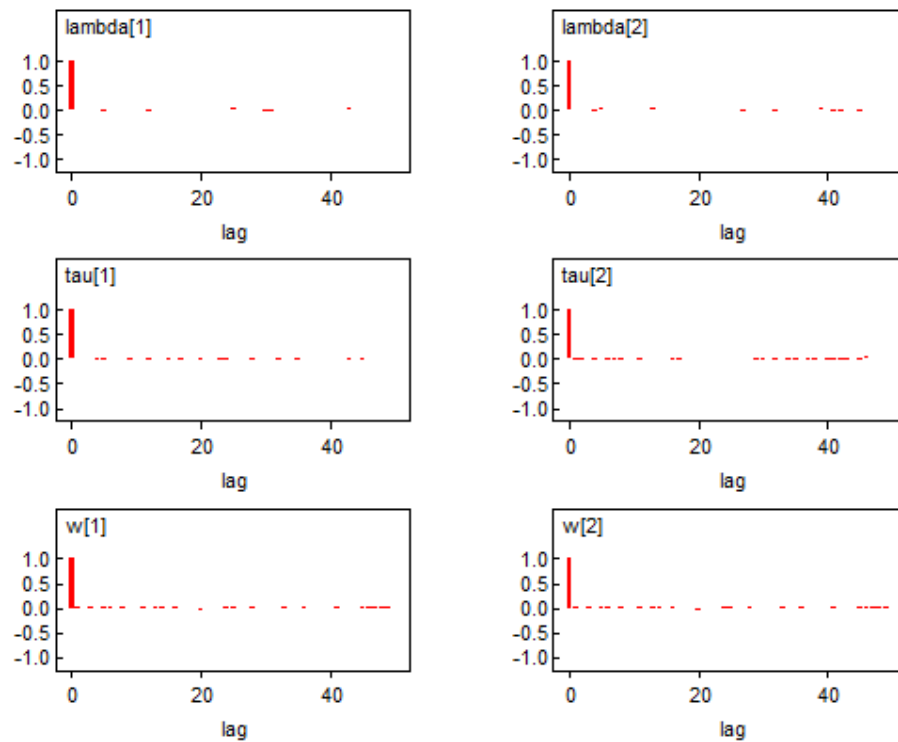
**Gambar L.4.59.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702

### 16. ID Gen CA1120 (Pembangkitan Parameter 5.000 Kali)

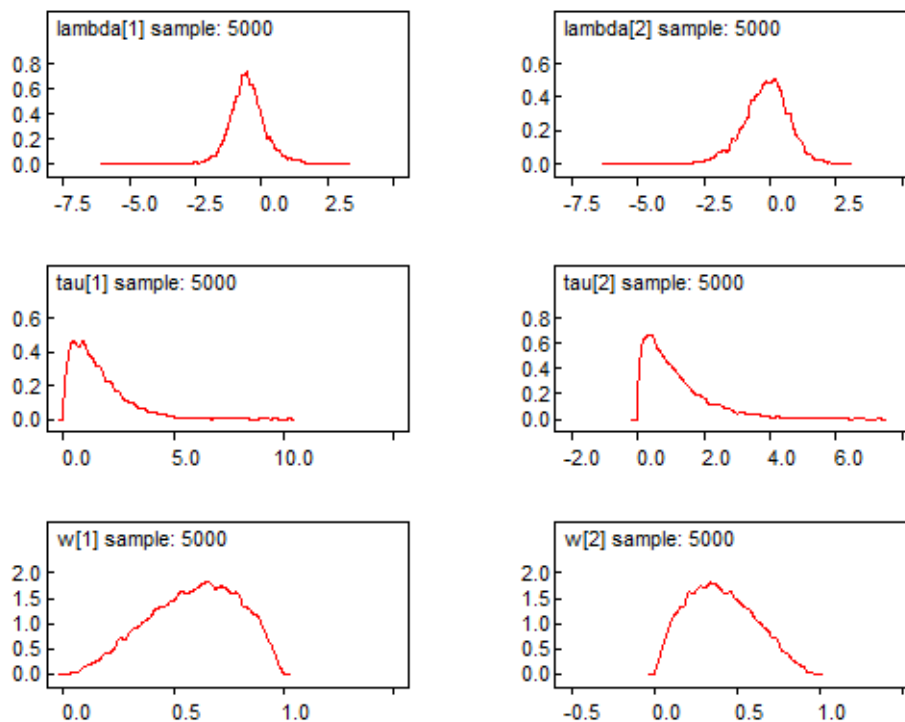
**Tabel L.4.6.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA1120

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5181	0,6869	0,0095	-1,811	-0,55	1,033	1	5.000
lambda[2]	-0,1464	0,8715	0,0109	-2,024	-0,0993	1,484	1	5.000
tau[1]	1,544	1,223	0,0183	0,1411	1,246	4,624	1	5.000
tau[2]	1,136	0,9938	0,014	0,0668	0,8628	3,753	1	5.000
w[1]	0,5957	0,2023	0,0031	0,1879	0,6122	0,9293	1	5.000
w[2]	0,4043	0,2023	0,0031	0,0711	0,3878	0,8122	1	5.000

### Lampiran 5. (Lanjutan)

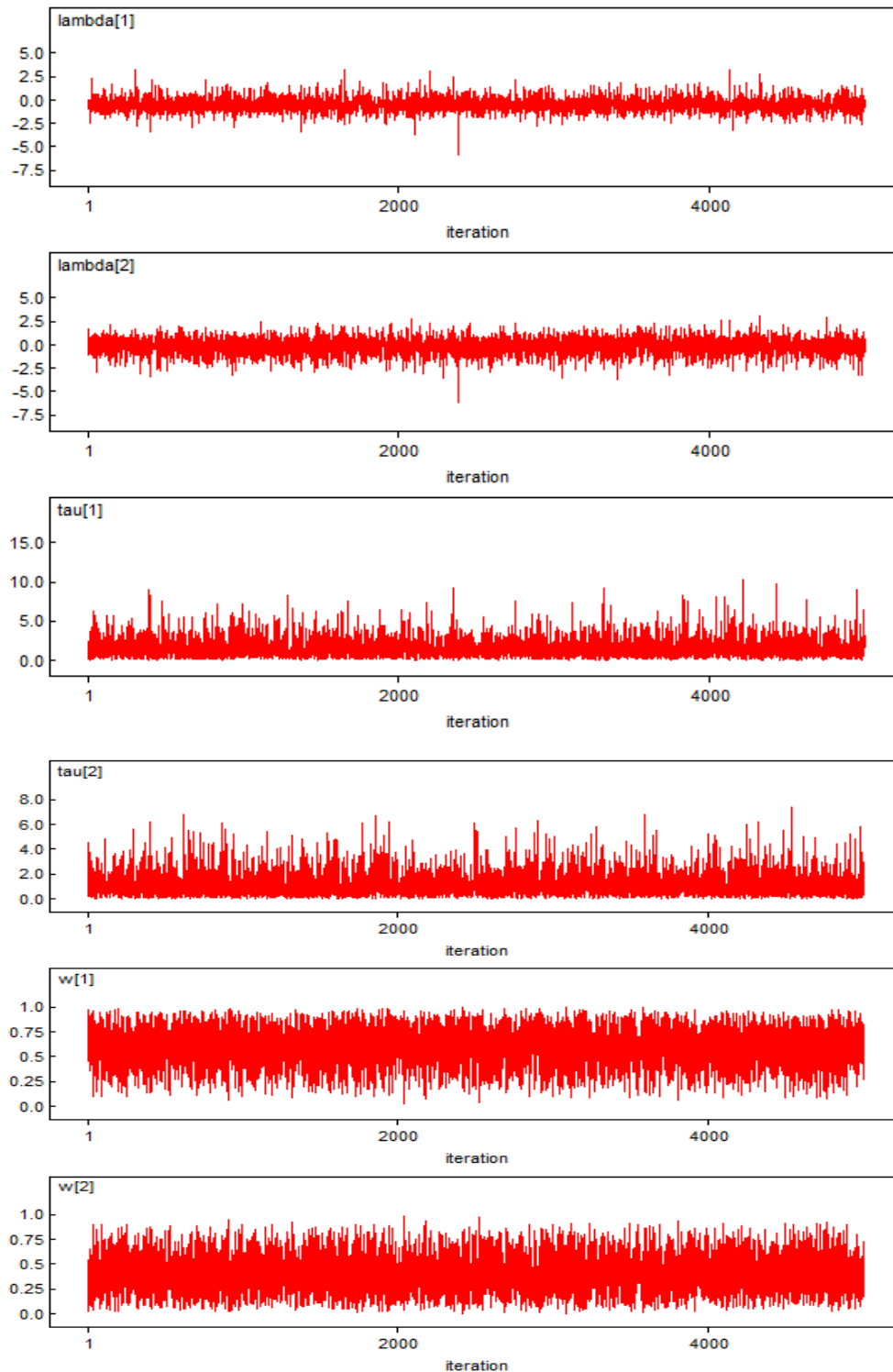


**Gambar L.4.60.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120



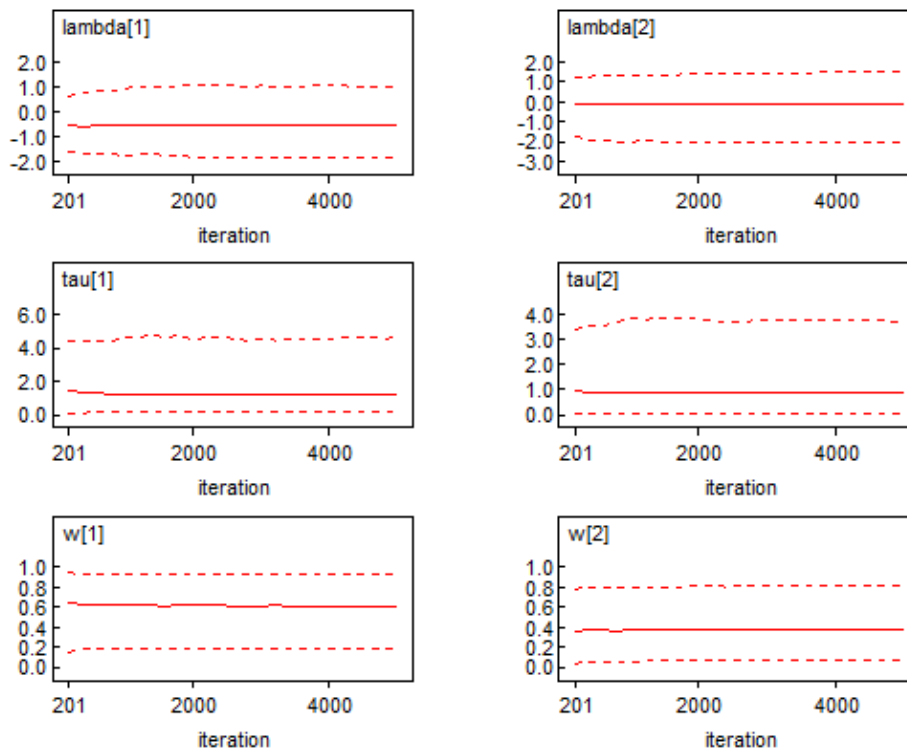
**Gambar L.4.61.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120

## Lampiran 5. (Lanjutan)



**Gambar L.4.62.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120

## Lampiran 5. (Lanjutan)



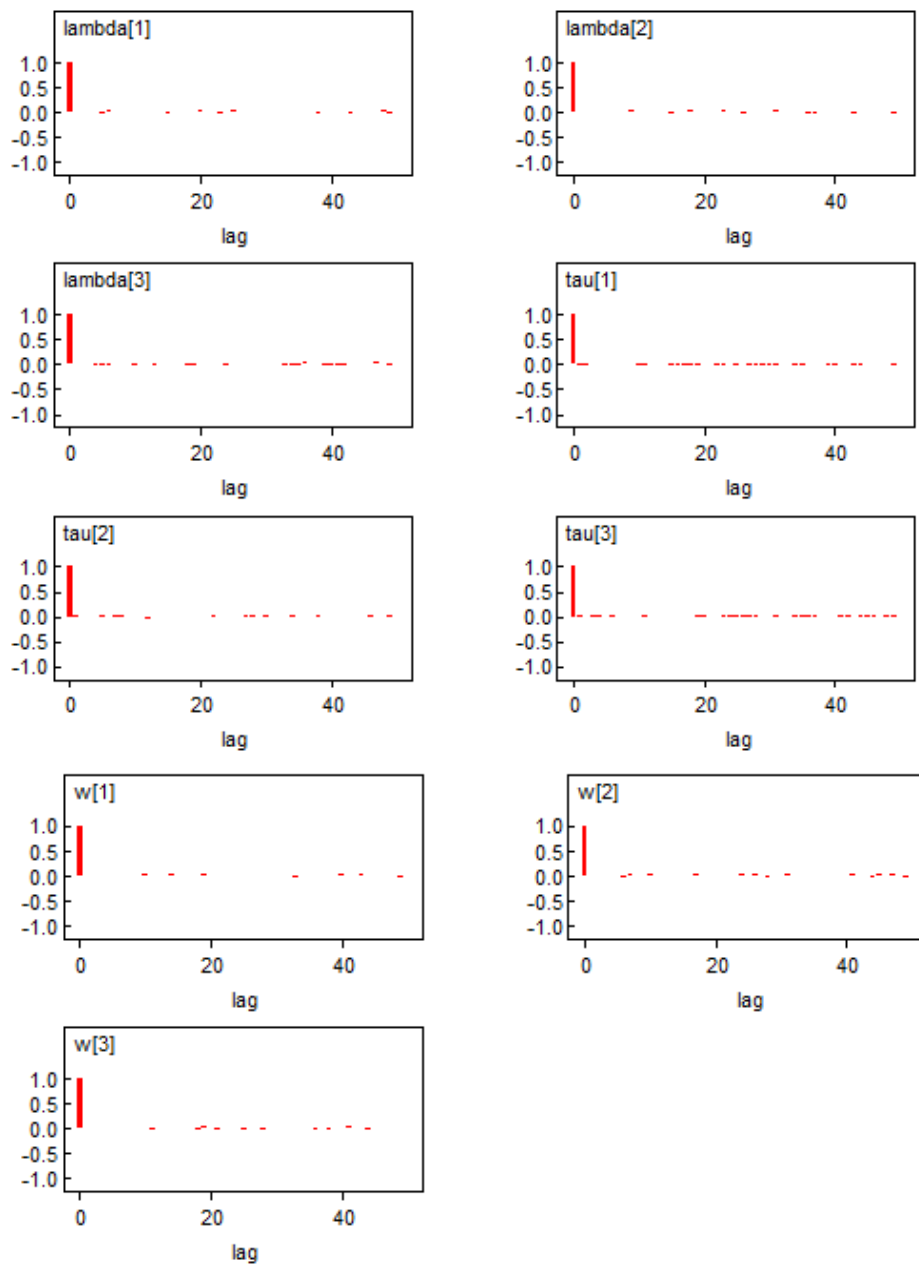
**Gambar L.4.63.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120

### 17. ID Gen U001 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.4.17.** Estimasi Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U001

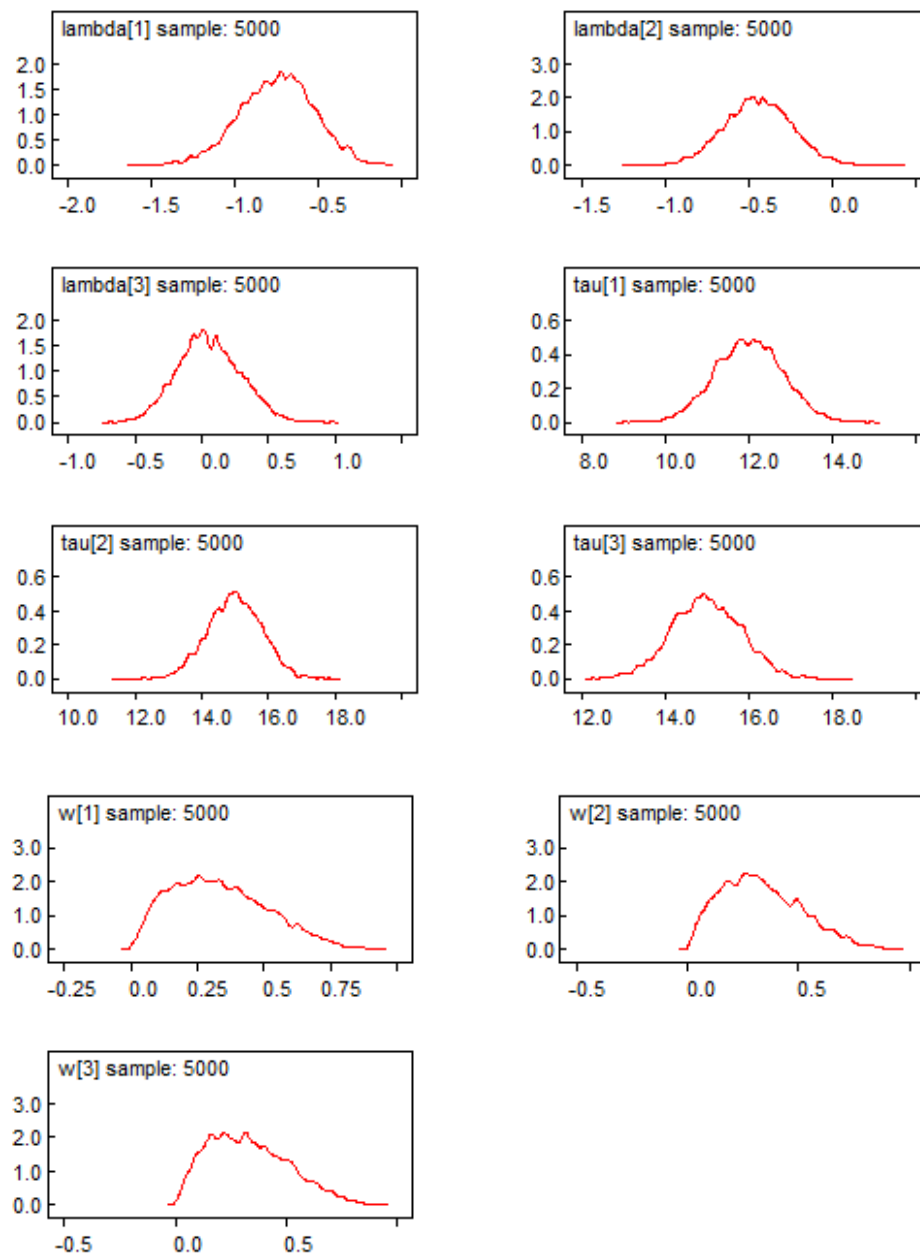
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,7487	0,2254	0,0036	-1,209	-0,7382	-0,3213	1	5.000
lambda[2]	-0,4397	0,202	0,0034	-0,833	-0,4405	-0,0376	1	5.000
lambda[3]	0,0557	0,2331	0,0034	-0,378	0,0439	0,5155	1	5.000
tau[1]	12,02	0,8061	0,0103	10,46	12,02	13,59	1	5.000
tau[2]	14,99	0,8114	0,0111	13,4	15	16,57	1	5.000
tau[3]	15	0,8286	0,0131	13,38	14,99	16,61	1	5.000
w[1]	0,3338	0,1777	0,0029	0,0544	0,3167	0,7068	1	5.000
w[2]	0,3334	0,1779	0,0025	0,0526	0,313	0,7193	1	5.000
w[3]	0,3328	0,1786	0,0026	0,0536	0,3154	0,7136	1	5.000

## Lampiran 5. (Lanjutan)



**Gambar L.4.64.** Fungsi Autokorelasi setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

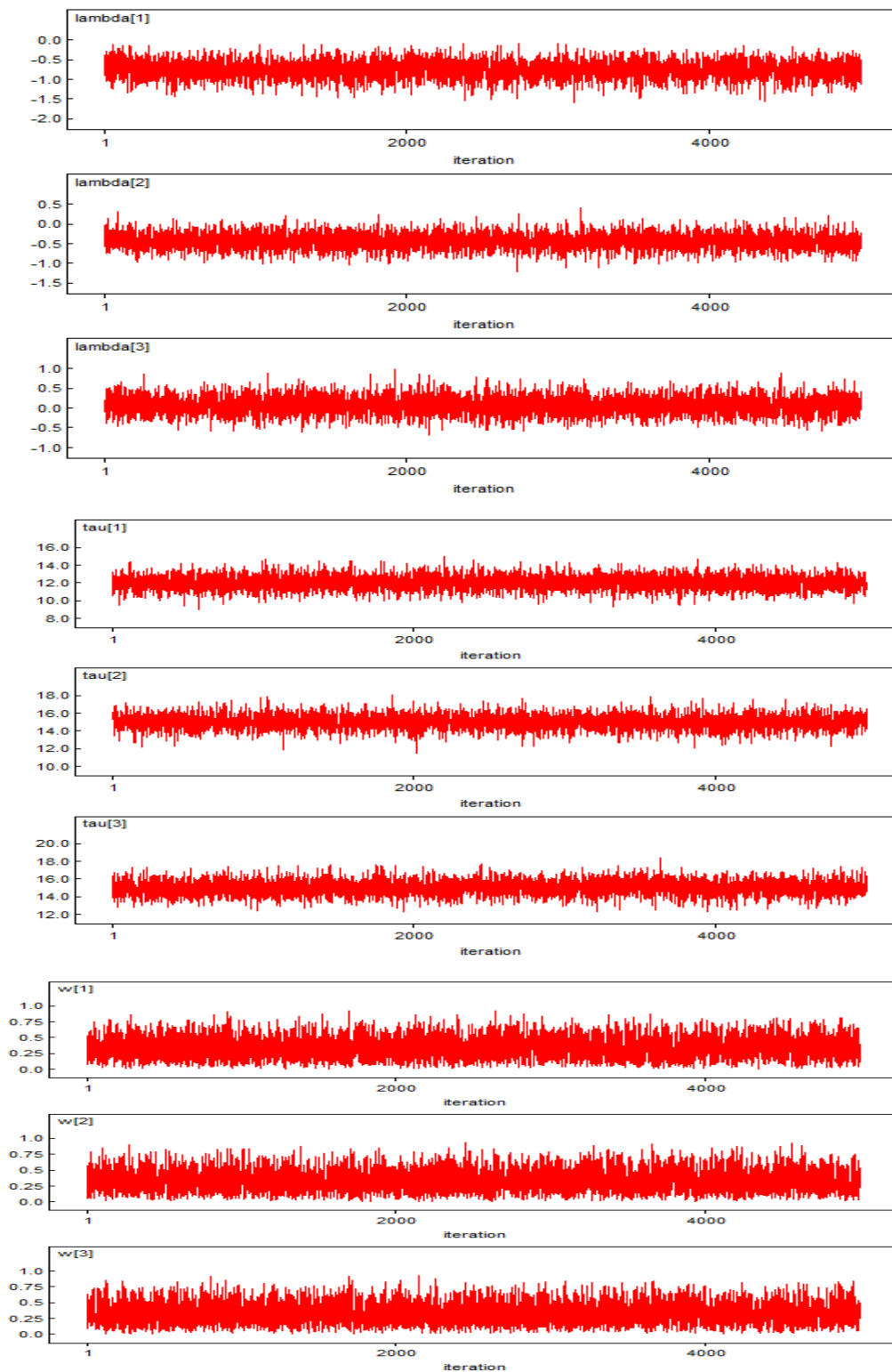
### Lampiran 5. (Lanjutan)



**Gambar L.4.65.** Densitas Kernel setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

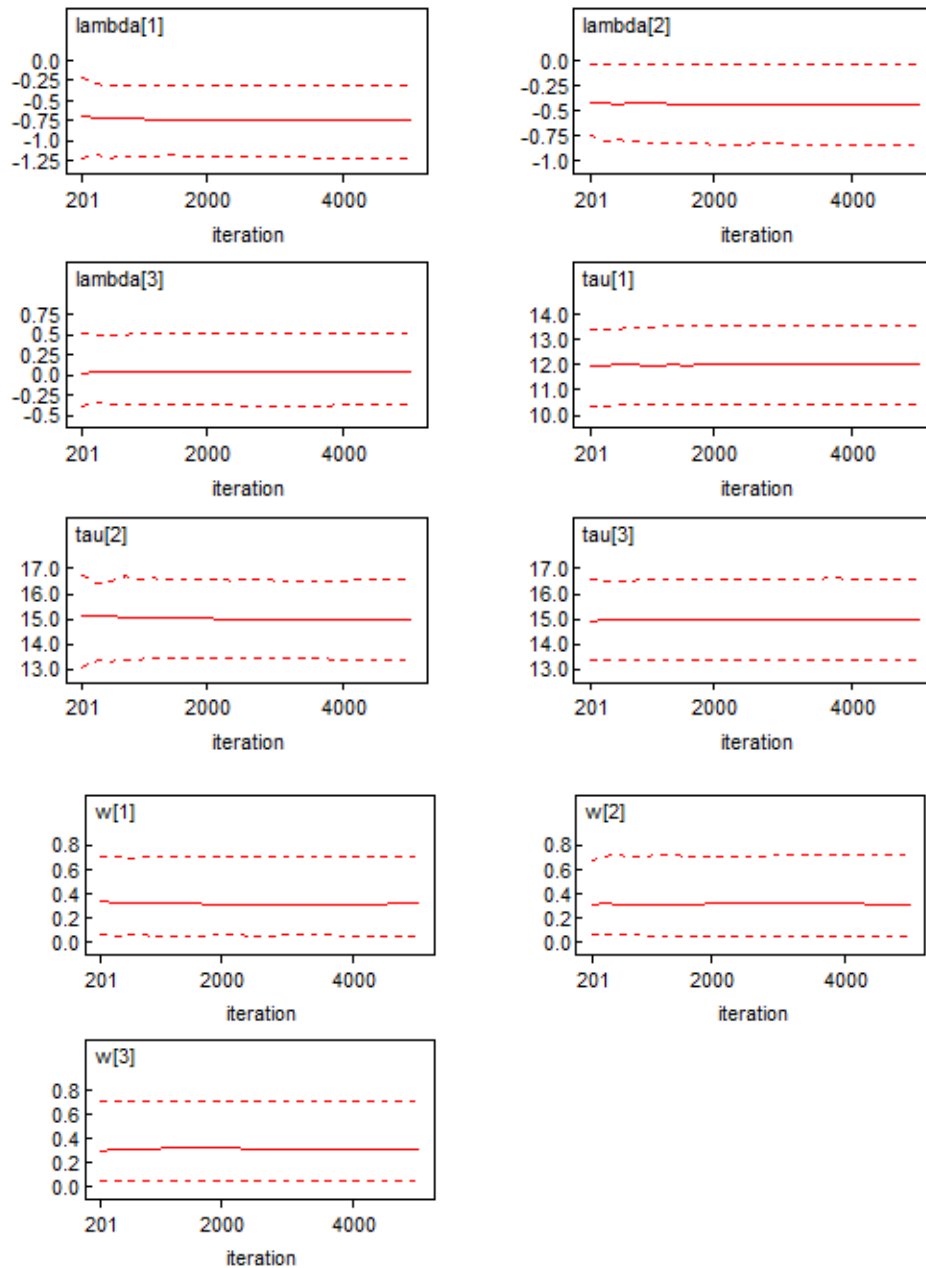


## Lampiran 5. (Lanjutan)



**Gambar L.4.66.** *History* Data Bangkitan setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

## Lampiran 5. (Lanjutan)



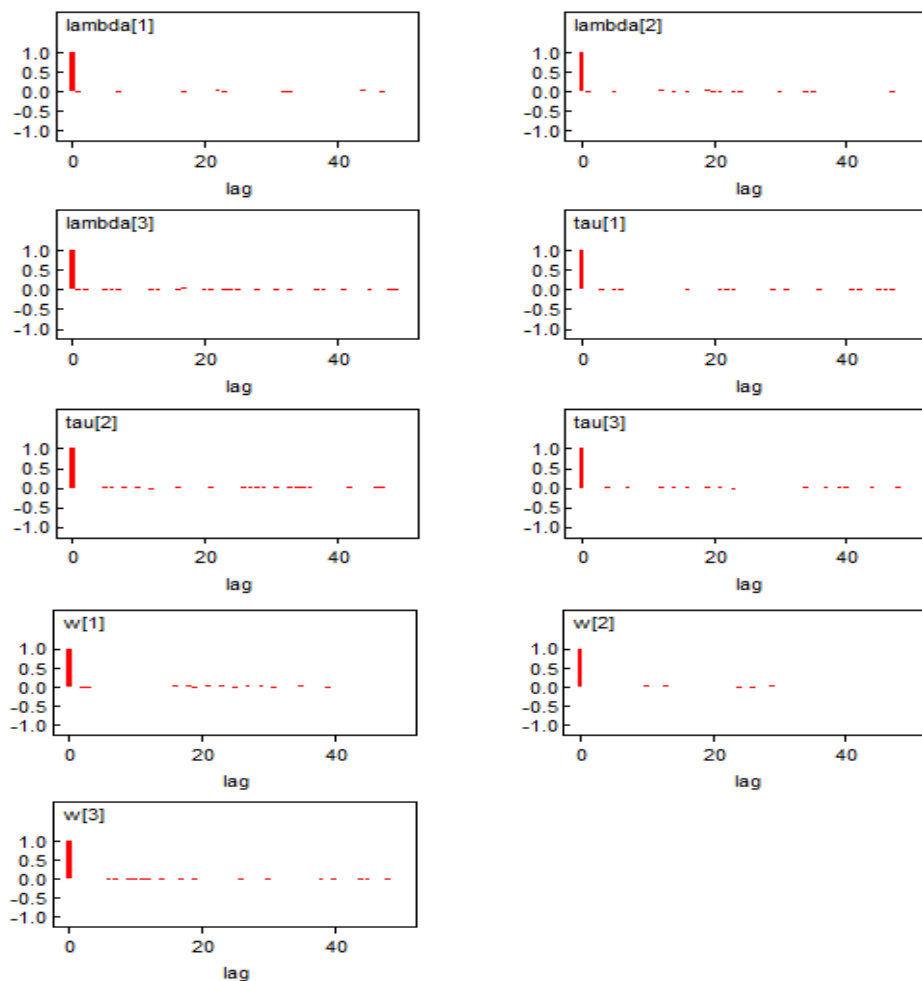
**Gambar L.4.67.** *Quantiles* setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

## Lampiran 5. (Lanjutan)

### 18. ID Gen CA0810 (Pembangkitan Parameter 5.000 Kali)

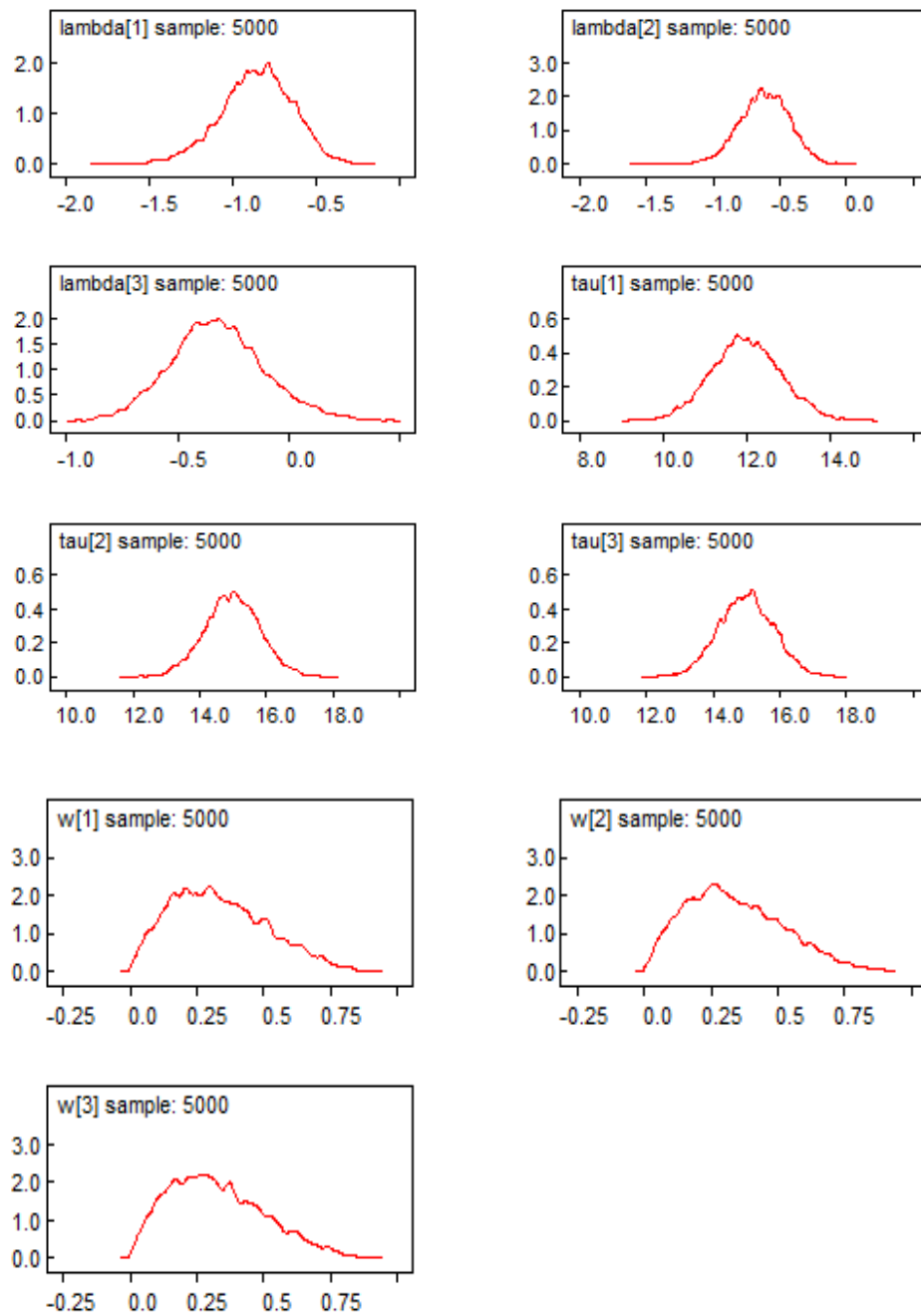
**Tabel L.4.18.** Estimasi Parameter Model BMM-Normal TigaKomponen  
untuk ID Gen CA0810

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,8572	0,2143	0,0029	-1,311	-0,849	-0,4723	1	5.000
lambda[2]	-0,6118	0,1844	0,0024	-0,981	-0,6101	-0,2616	1	5.000
lambda[3]	-0,3188	0,2088	0,0029	-0,706	-0,3264	0,1258	1	5.000
tau[1]	12	0,823	0,0116	10,39	11,99	13,62	1	5.000
tau[2]	14,98	0,8194	0,0108	13,35	14,99	16,62	1	5.000
tau[3]	15	0,8251	0,0106	13,41	15	16,61	1	5.000
w[1]	0,3339	0,177	0,0022	0,0511	0,3133	0,7135	1	5.000
w[2]	0,3399	0,1788	0,0024	0,0573	0,3182	0,7238	1	5.000
w[3]	0,3262	0,1762	0,0022	0,0505	0,305	0,708	1	5.000



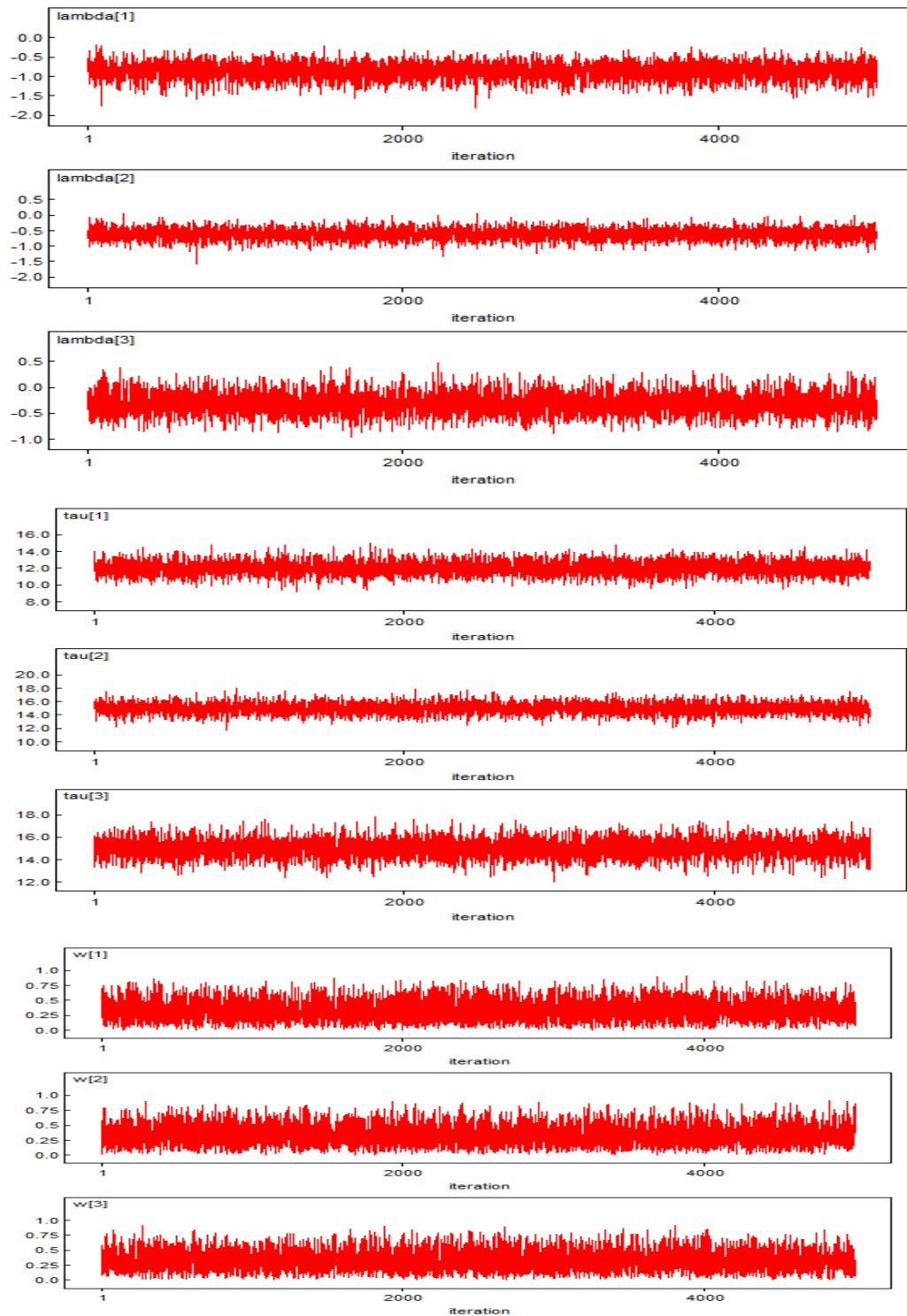
**Gambar L.4.68.** Fungsi Autokorelasi setiap Parameter BMM-Normal Tiga  
Komponen untuk ID Gen CA0810

### Lampiran 5. (Lanjutan)



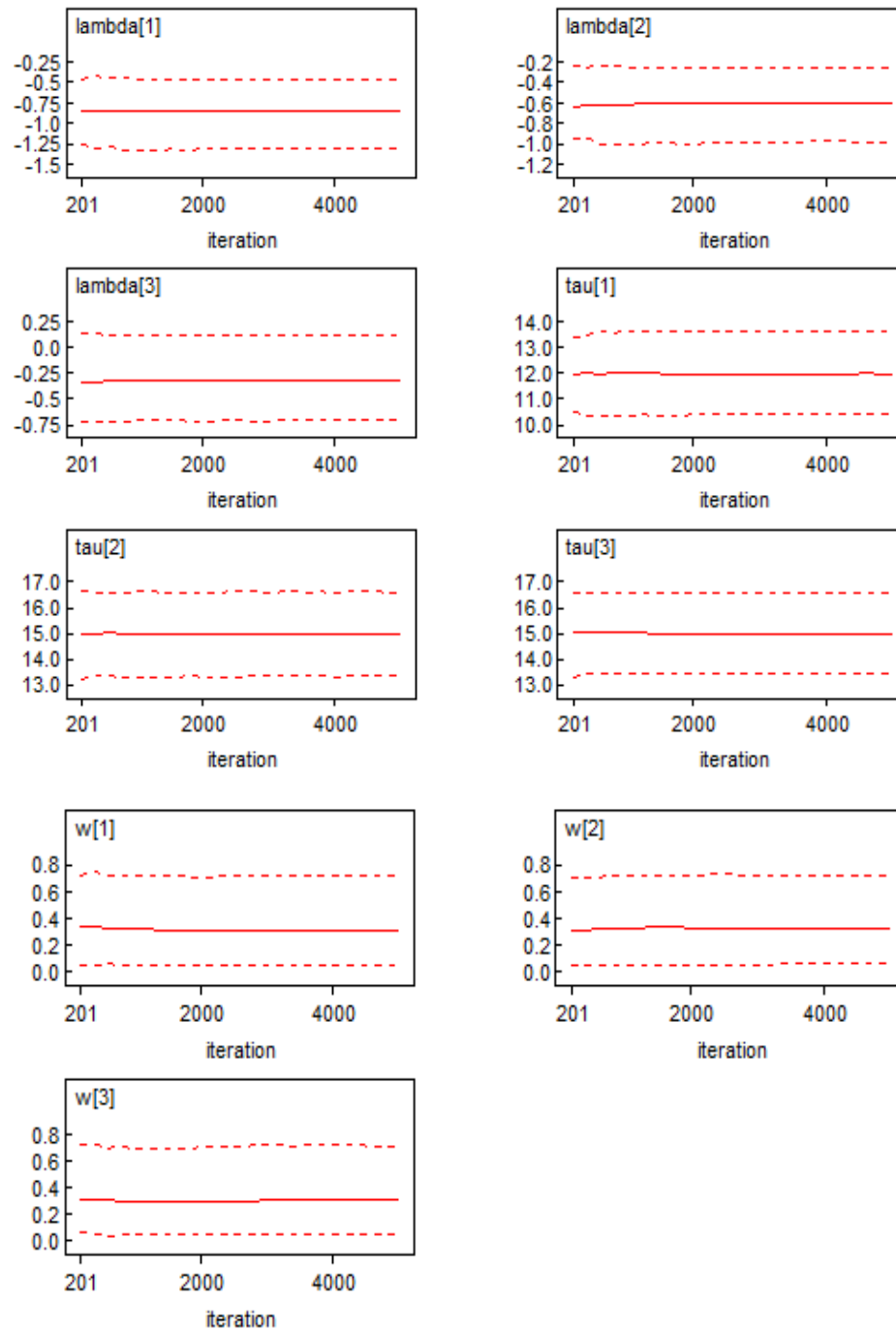
**Gambar L.4.69.** Densitas Kernel setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen CA0810

## Lampiran 5. (Lanjutan)



**Gambar L.4.70.** *History* Data Bangkitan setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen CA0810

## Lampiran 5. (Lanjutan)



**Gambar L.4.71.** *Quantiles* setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen CA0810

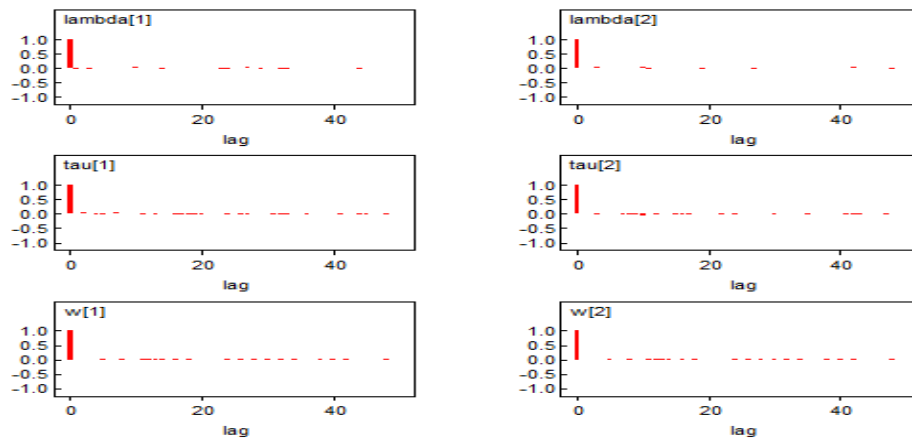
*Halaman ini sengaja dikosongkan*

**Lampiran 6.** Hasil *Running* Program BUGS untuk 19 ID Gen pada Model BMM-Normal Data *Chickpea*

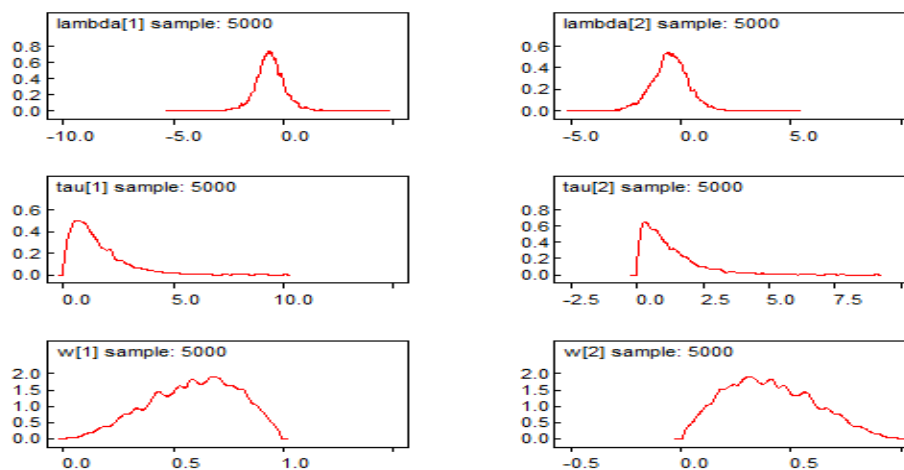
1. ID Gen LS0616 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.5.1.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0616

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5935	0,6555	0,0081	-1,907	-0,6108	0,7444	1	5.000
lambda[2]	-0,4969	0,8296	0,0111	-2,196	-0,4877	1,136	1	5.000
tau[1]	1,481	1,144	0,0164	0,1484	1,183	4,443	1	5.000
tau[2]	1,205	1,041	0,0135	0,0833	0,913	4,065	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000



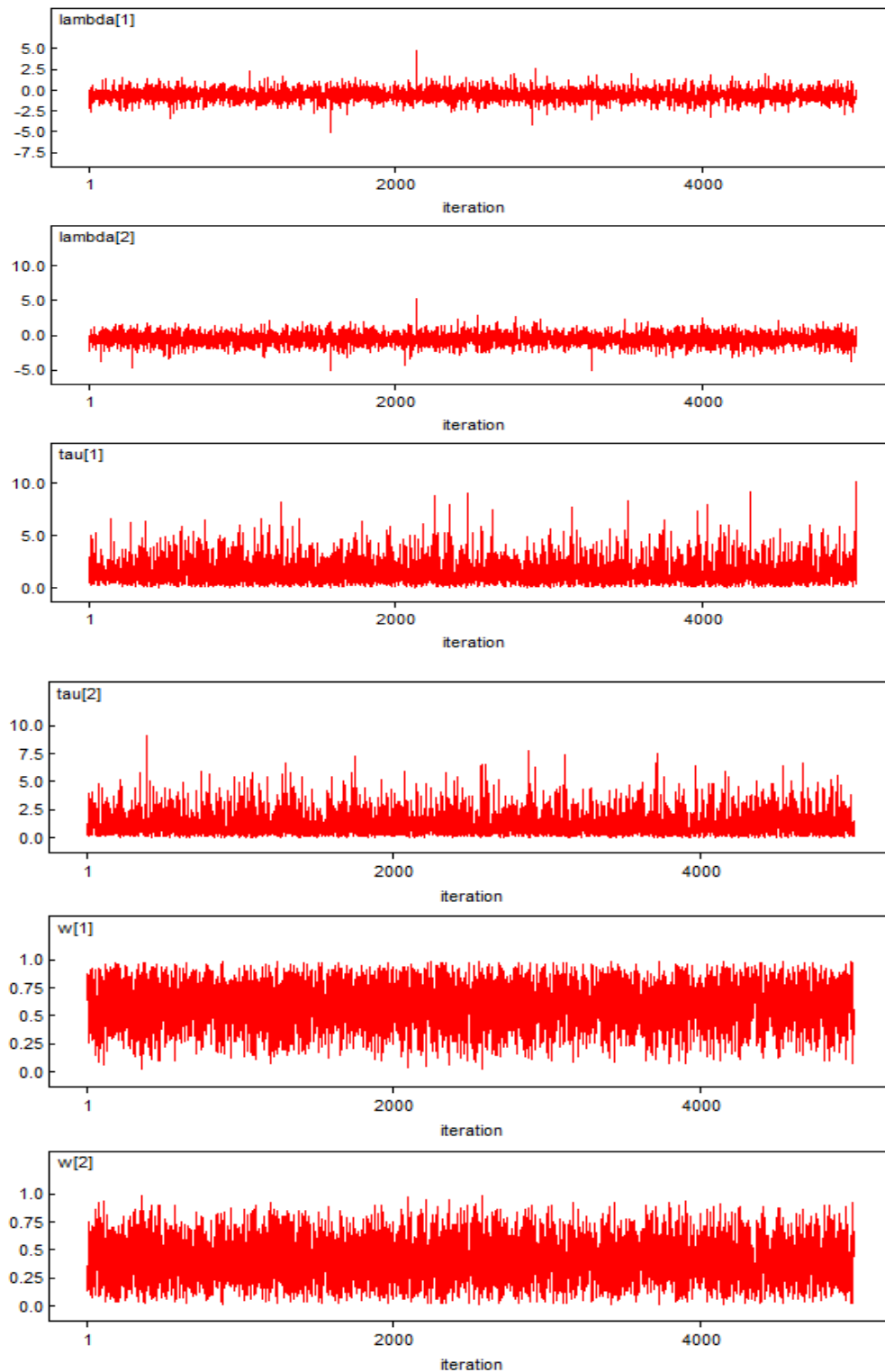
**Gambar L.5.1.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616



**Gambar L.5.2.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616

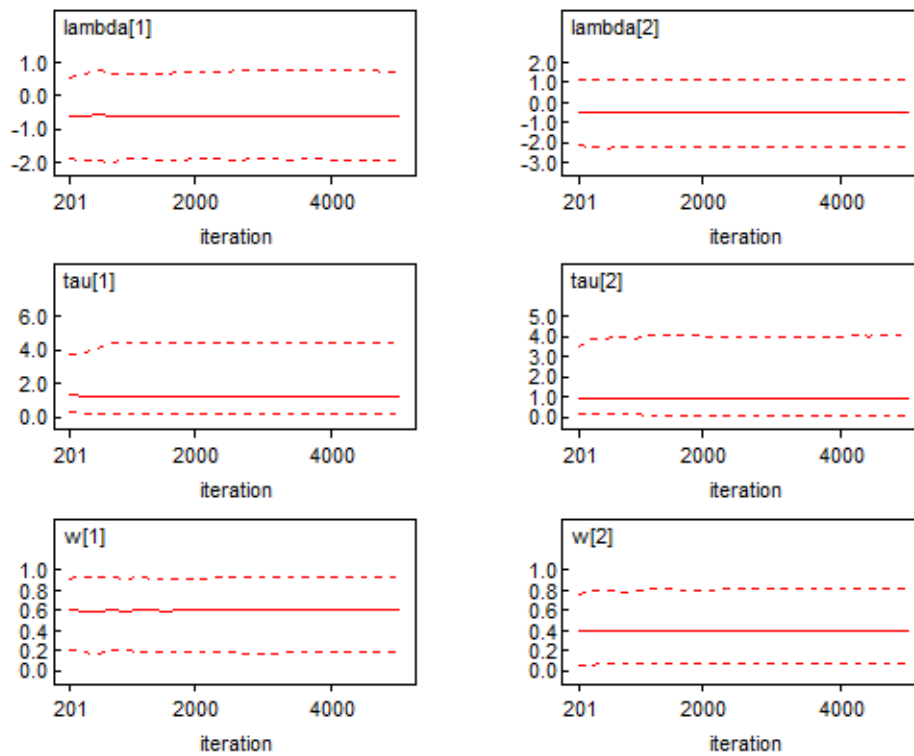


## Lampiran 6. (Lanjutan)



**Gambar L.5.3.** *History Data* Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616

## Lampiran 6. (Lanjutan)



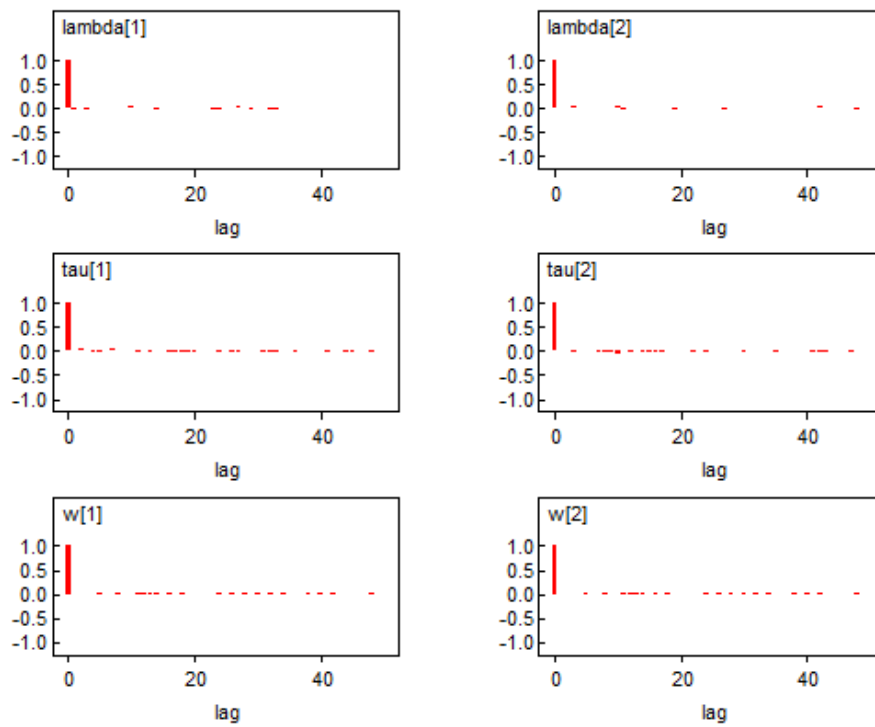
**Gambar L.5.4.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0616

## 2. ID Gen LS0159 (Pembangkitan Parameter 5.000 Kali)

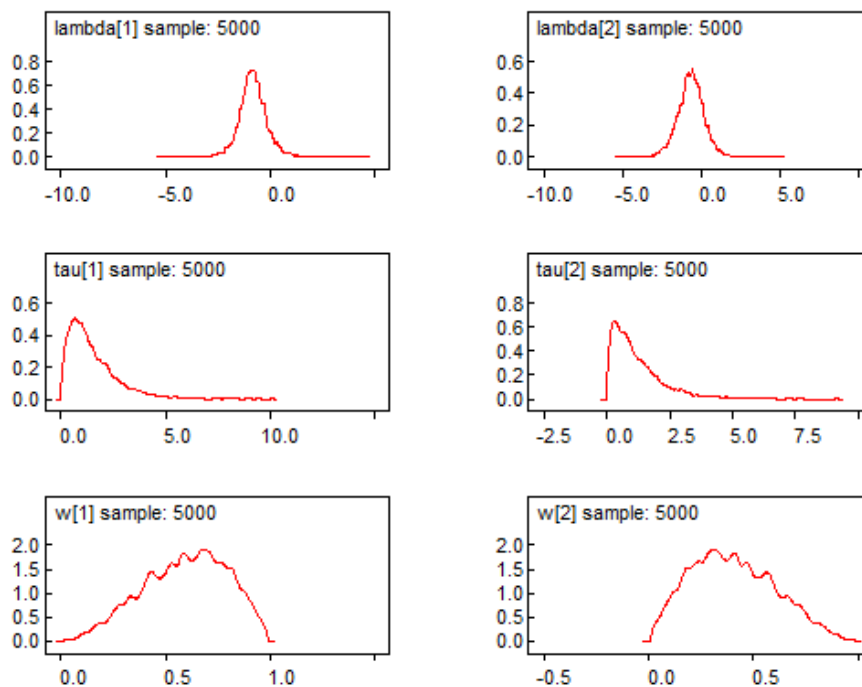
**Tabel L.5.2.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0159

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,8085	0,6561	0,008	-2,102	-0,8324	0,5563	1	5.000
lambda[2]	-0,6589	0,8335	0,0111	-2,386	-0,6408	0,9575	1	5.000
tau[1]	1,497	1,16	0,0168	0,1492	1,201	4,491	1	5.000
tau[2]	1,199	1,039	0,0134	0,0821	0,9064	4,043	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

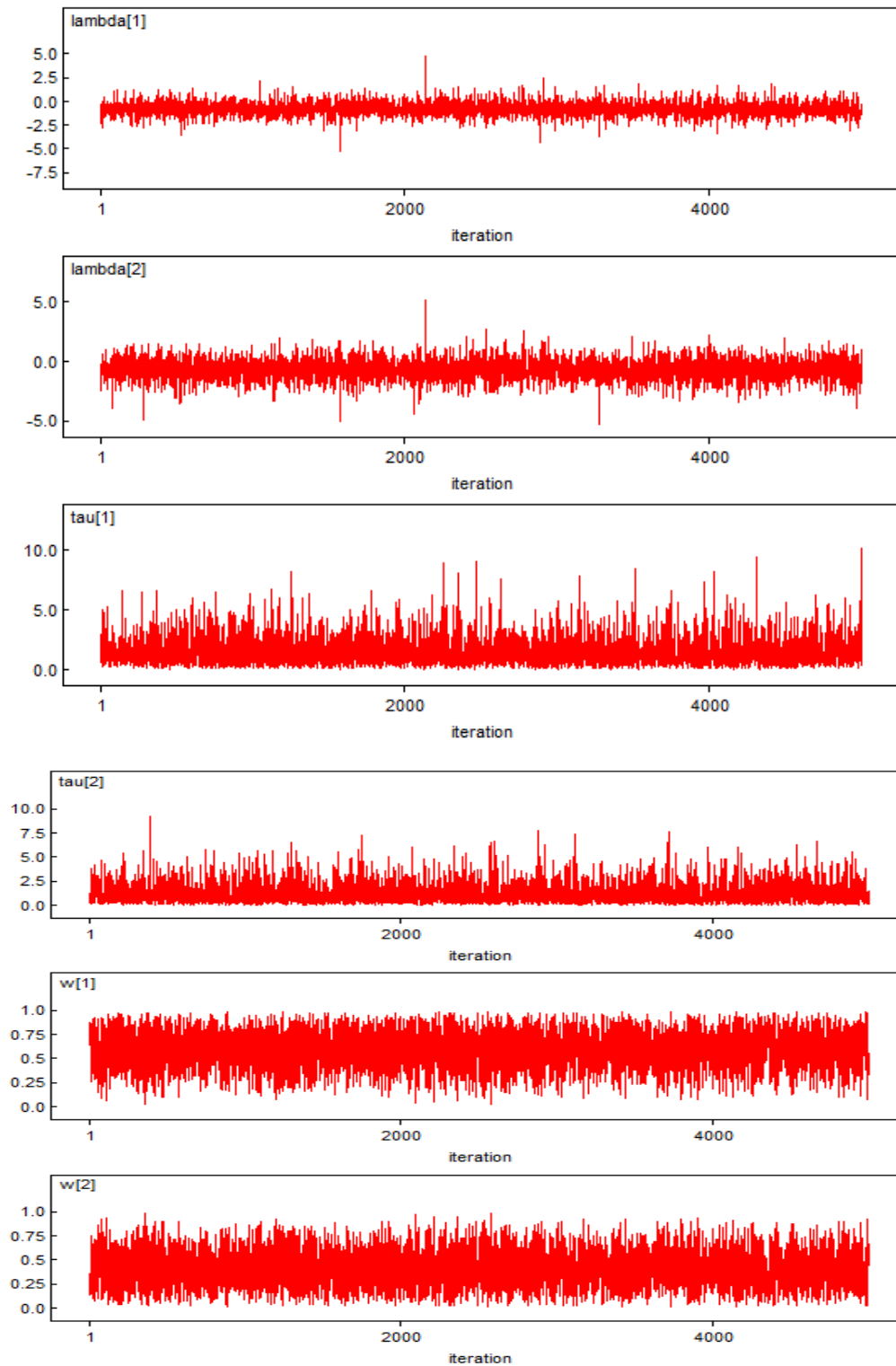


**Gambar L.5.5.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159



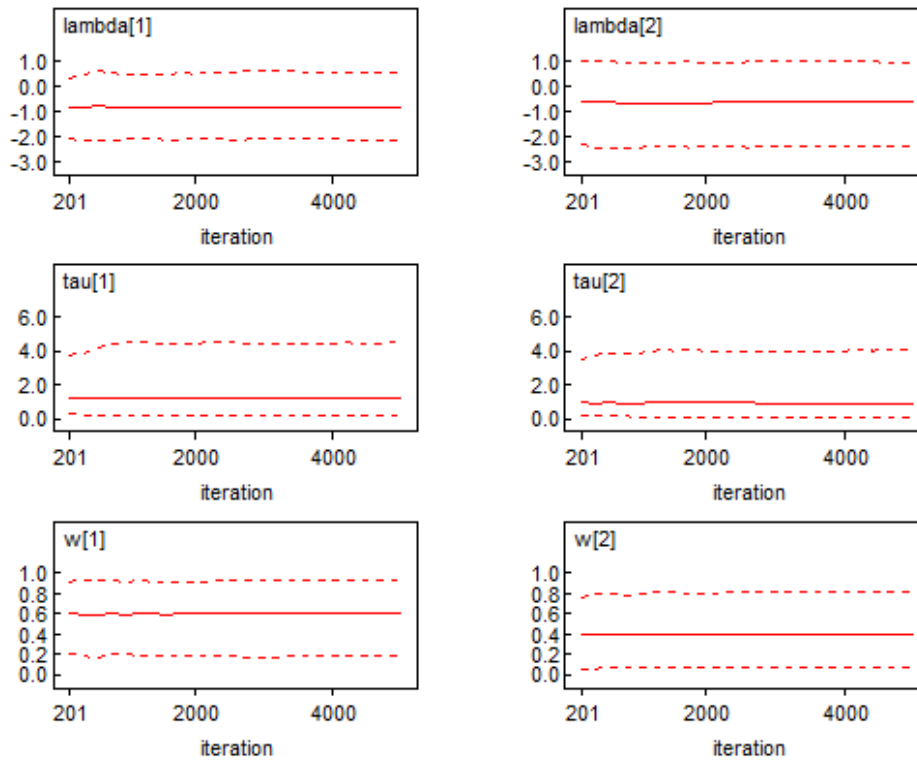
**Gambar L.5.6.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159

## Lampiran 6. (Lanjutan)



**Gambar L.5.7.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159*

## Lampiran 6. (Lanjutan)



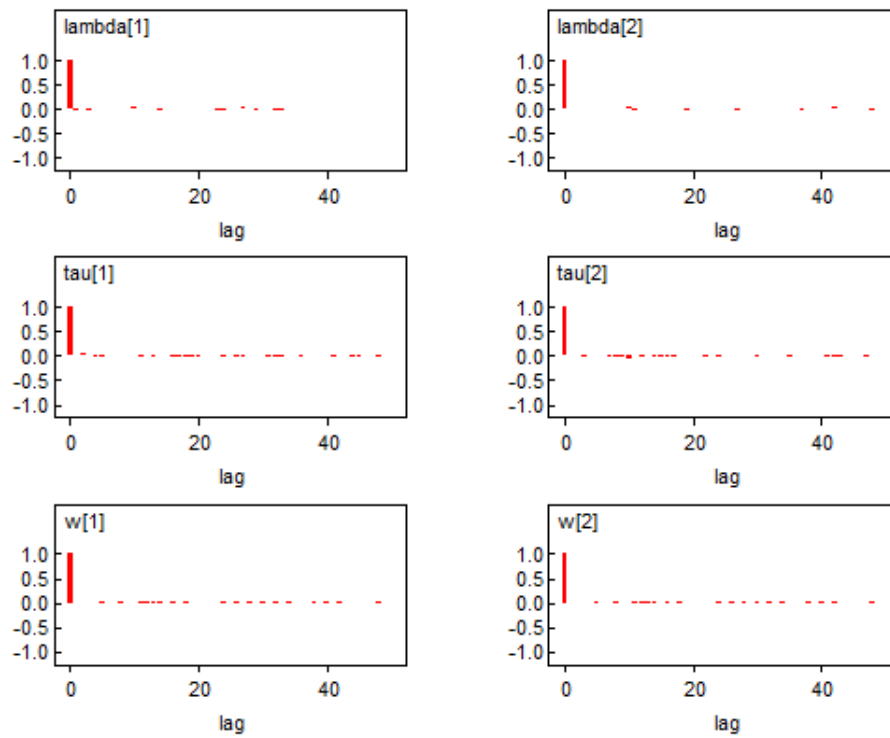
**Gambar L.5.8.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0159

### 3. ID Gen LS0162 (Pembangkitan Parameter 5.000 Kali)

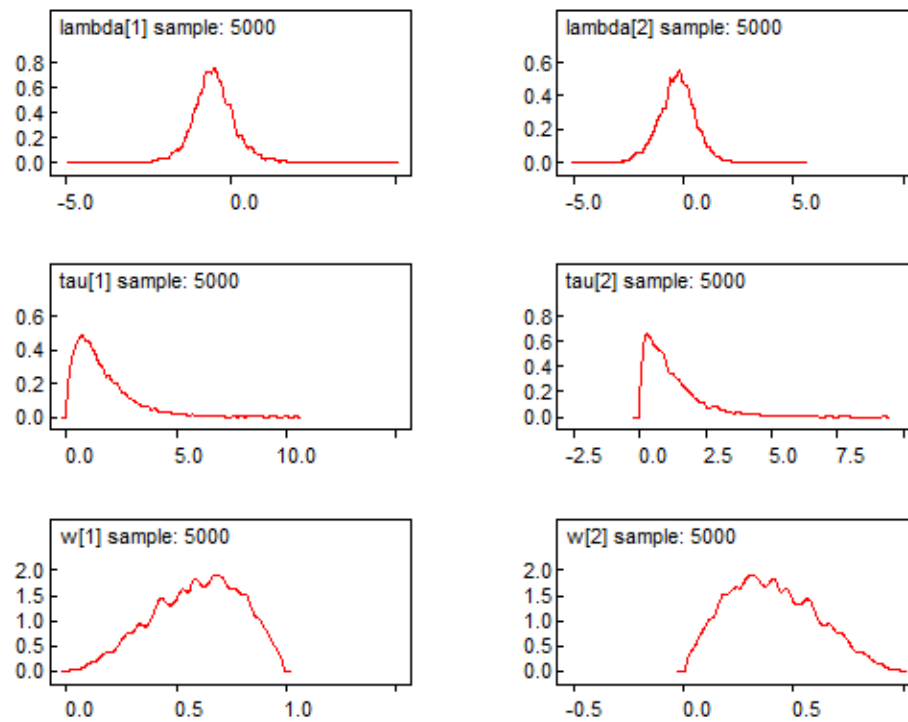
**Tabel L.5.3.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0162

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5194	0,649	0,0079	-1,779	-0,5421	0,8649	1	5.000
lambda[2]	-0,304	0,838	0,0112	-2,062	-0,2769	1,309	1	5.000
tau[1]	1,572	1,222	0,0178	0,1546	1,256	4,716	1	5.000
tau[2]	1,189	1,036	0,0134	0,0823	0,8981	3,996	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

### Lampiran 6. (Lanjutan)

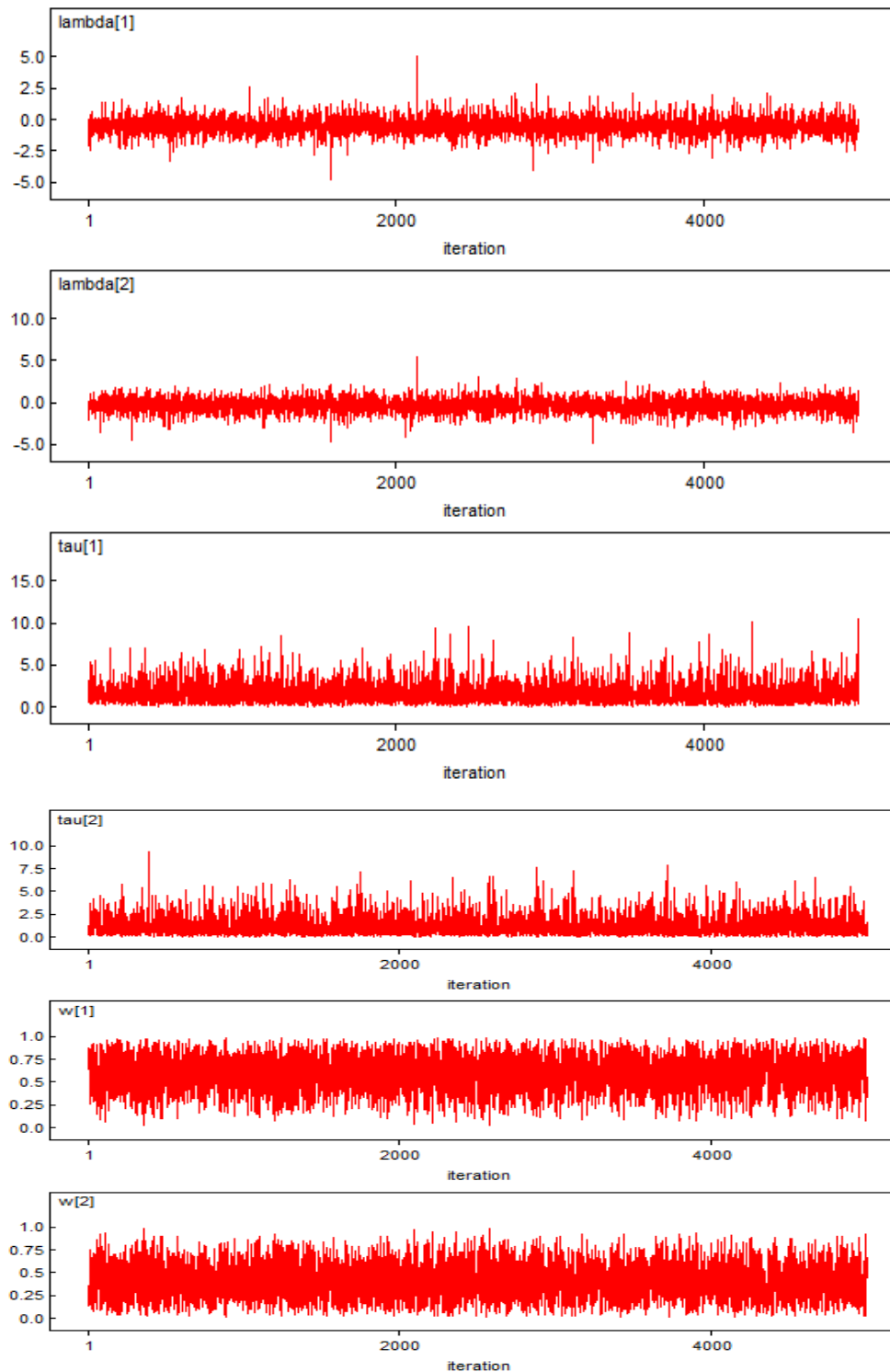


**Gambar L.5.9.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162



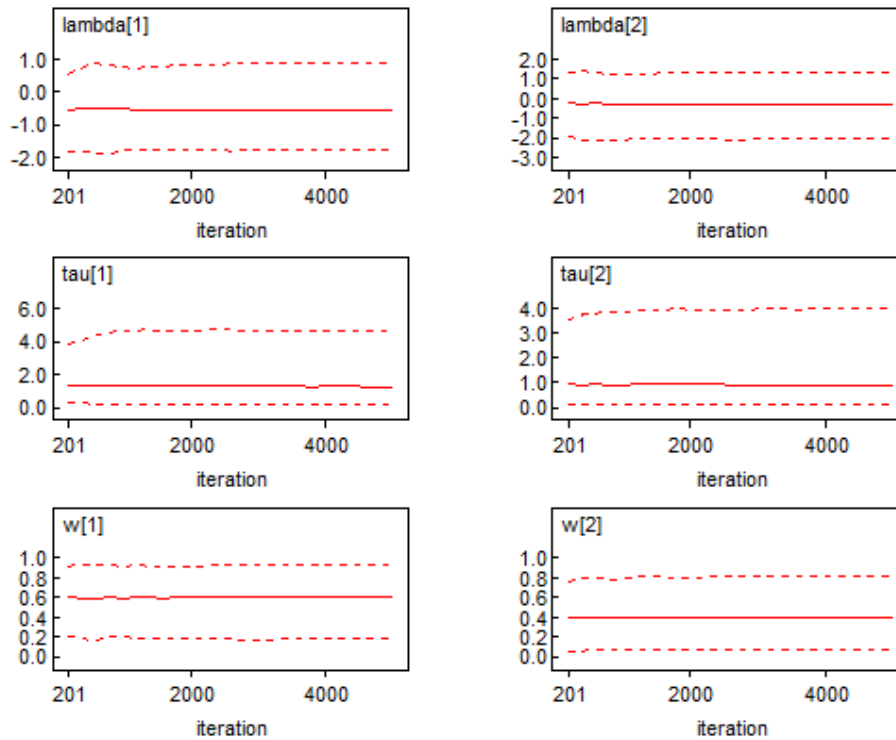
**Gambar L.5.10.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162

## Lampiran 6. (Lanjutan)



**Gambar L.5.11.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162*

## Lampiran 6. (Lanjutan)



**Gambar L.5.12.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0162

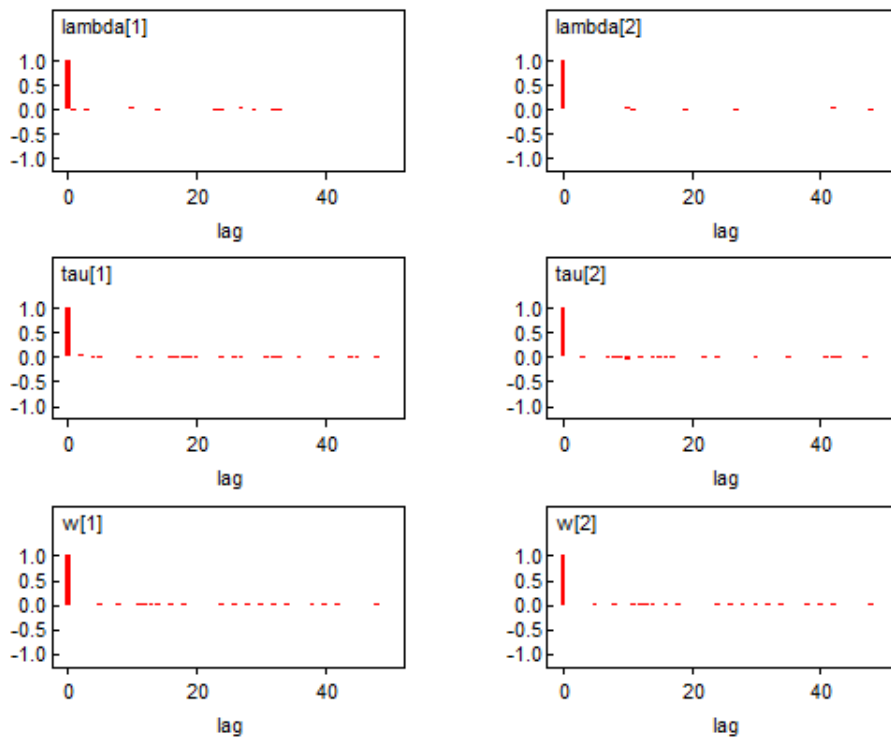
### 4. ID Gen LS0185 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.5.4.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0185

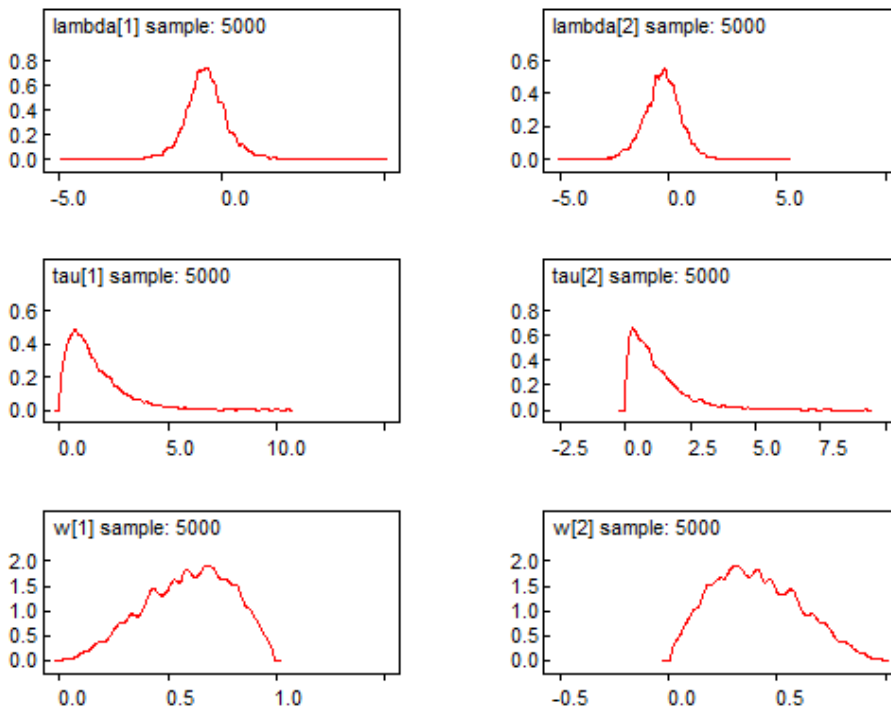
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,4813	0,6464	0,0079	-1,743	-0,5022	0,8872	1	5.000
lambda[2]	-0,2856	0,8349	0,0111	-2,033	-0,262	1,328	1	5.000
tau[1]	1,579	1,226	0,0178	0,1565	1,262	4,737	1	5.000
tau[2]	1,193	1,038	0,0134	0,0821	0,9011	4,012	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000



### Lampiran 6. (Lanjutan)

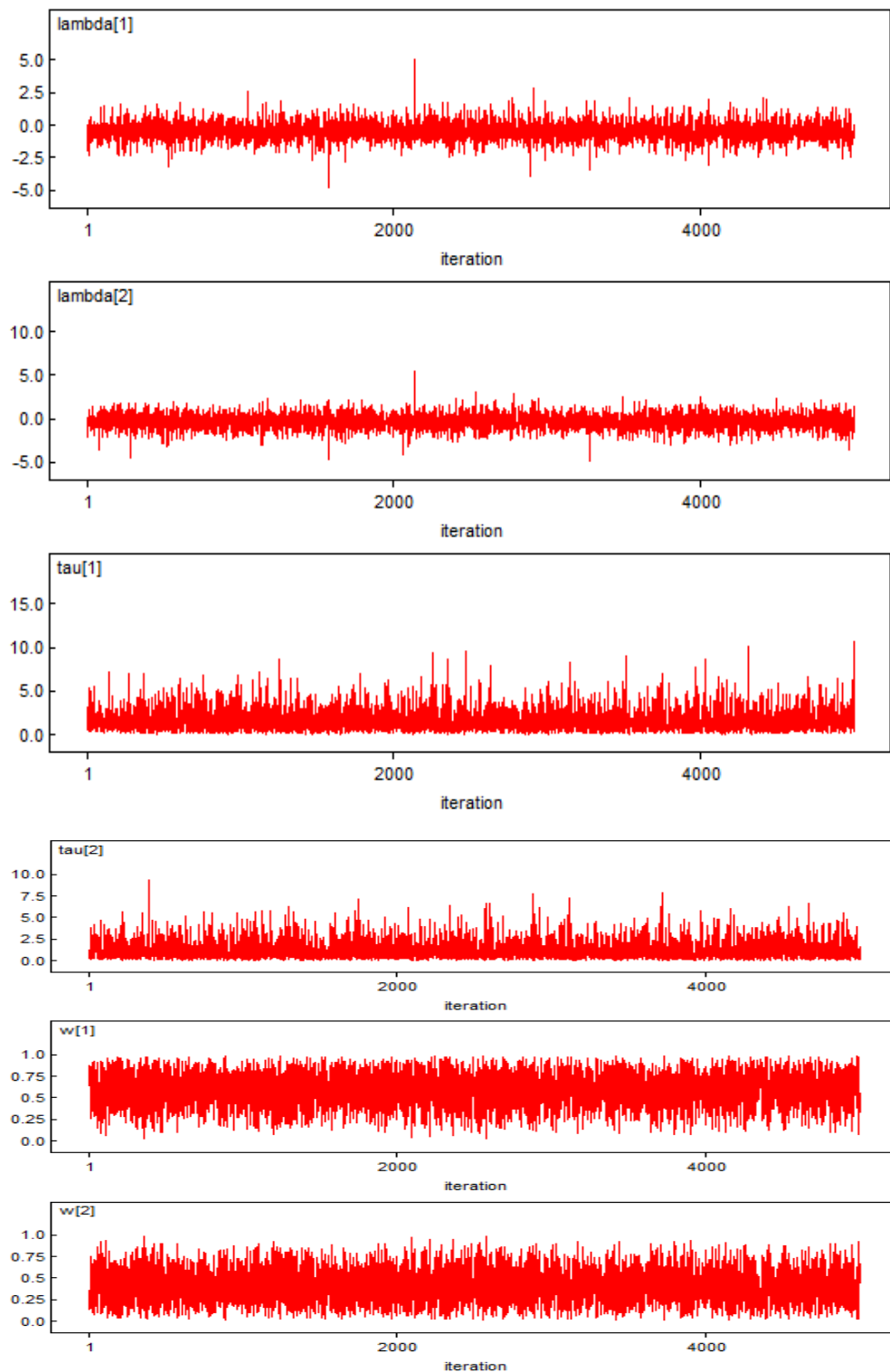


**Gambar L.5.13.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0185



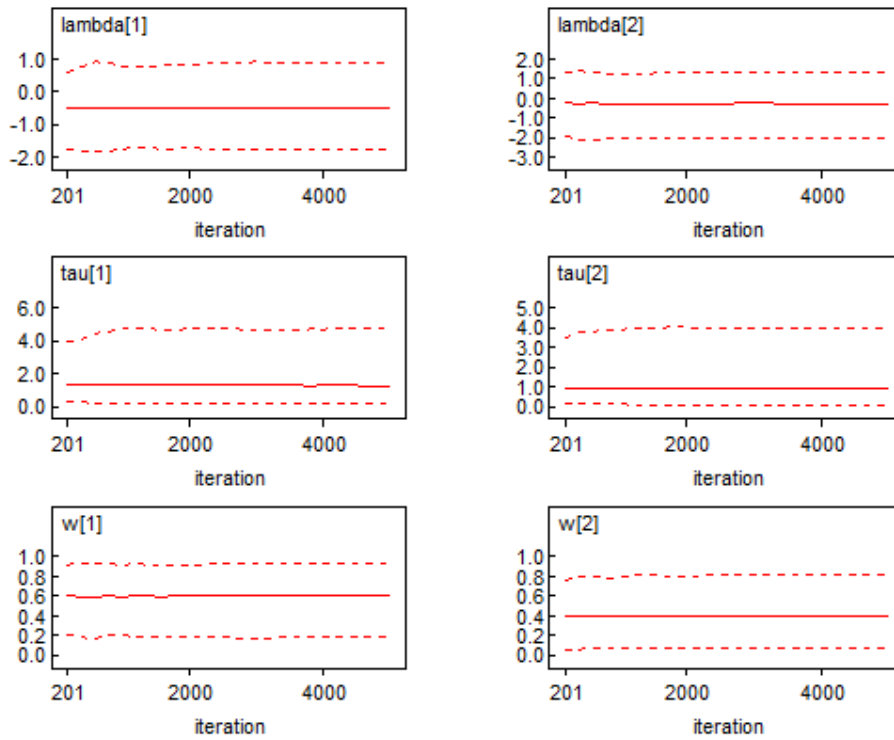
**Gambar L.5.14.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0185

## Lampiran 6. (Lanjutan)



**Gambar L.5.15.** *History Data* Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0185

## Lampiran 6. (Lanjutan)



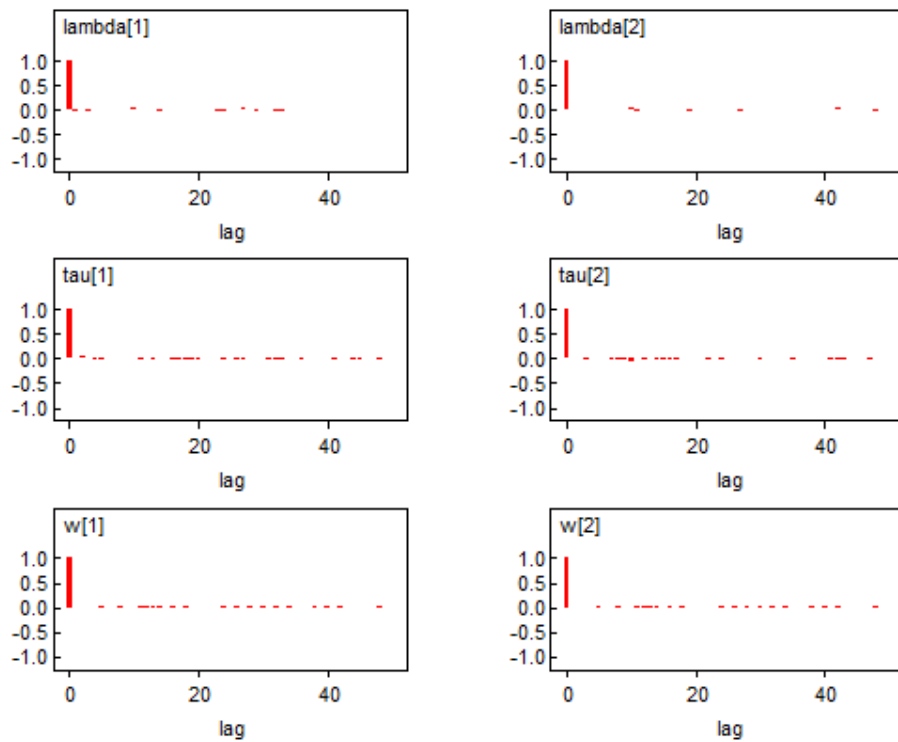
**Gambar L.5.16.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0185

## 5. ID Gen LS0081 (Pembangkitan Parameter 5.000 Kali)

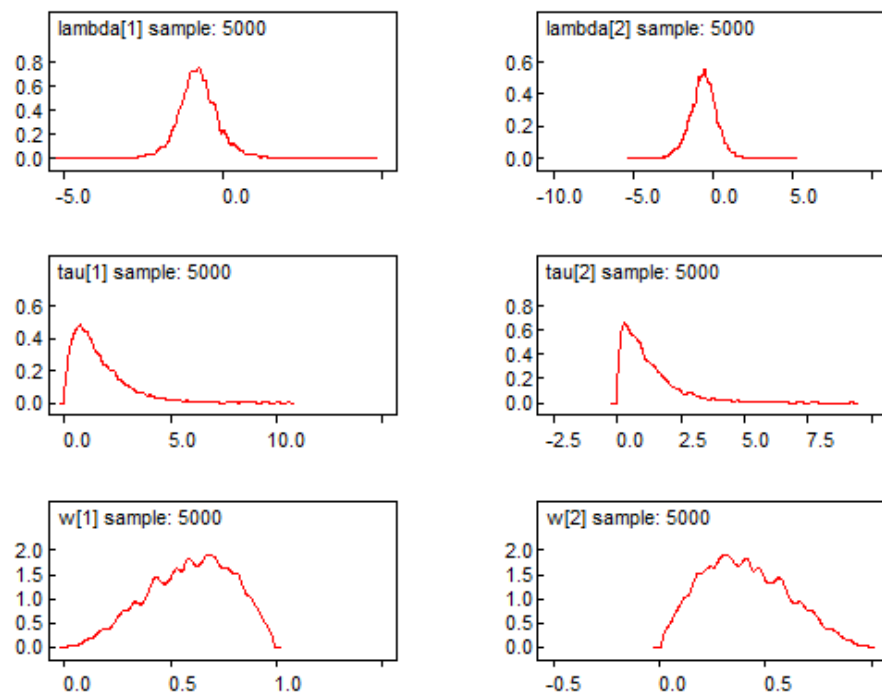
**Tabel L.5.5.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0081

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,7817	0,646	0,0079	-2,039	-0,8042	0,5971	1	5.000
lambda[2]	-0,587	0,8347	0,0111	-2,337	-0,5632	1,025	1	5.000
tau[1]	1,588	1,234	0,0179	0,1563	1,27	4,765	1	5.000
tau[2]	1,193	1,038	0,0134	0,0824	0,901	4,012	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

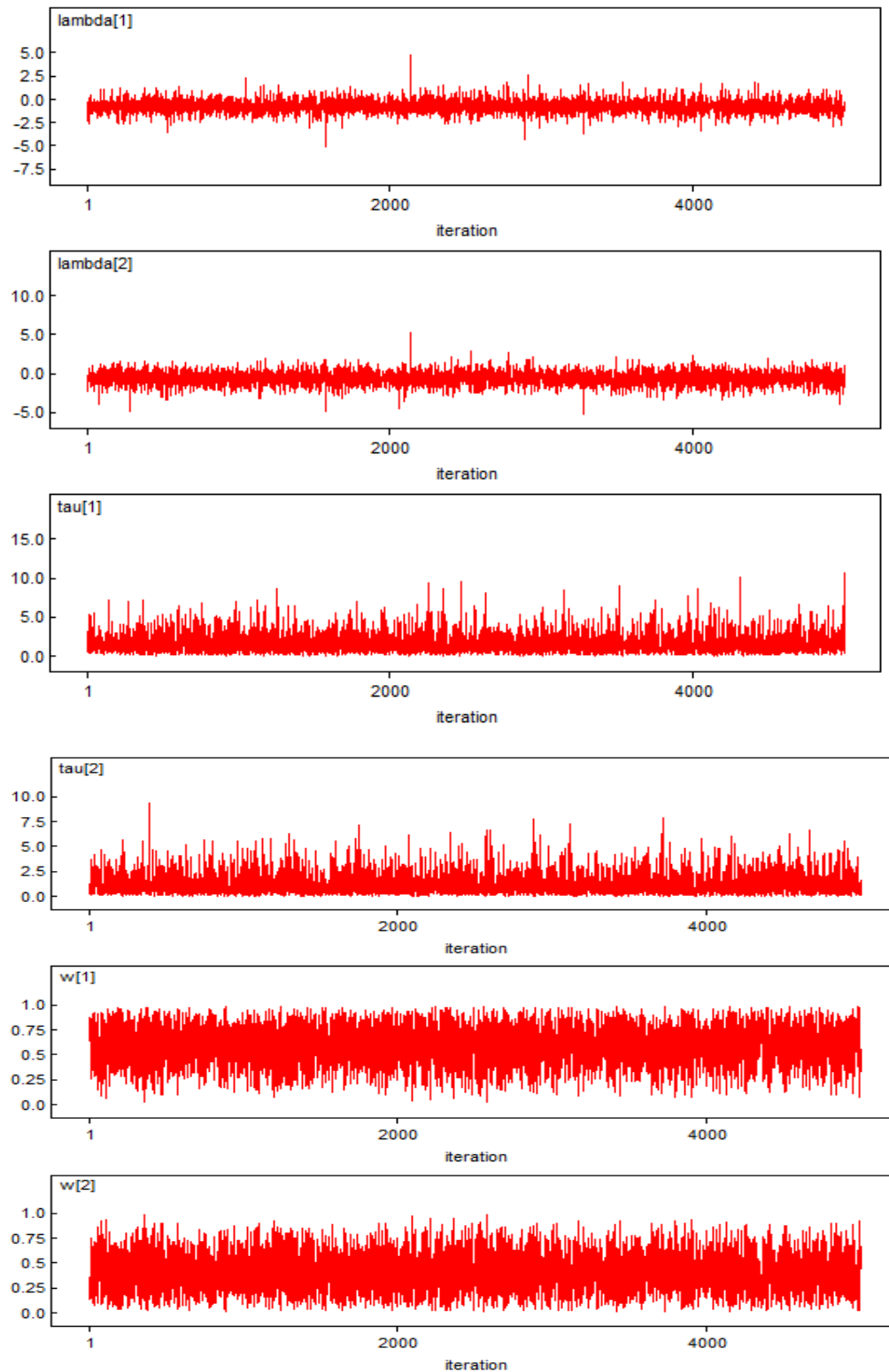


**Gambar L.5.17.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081



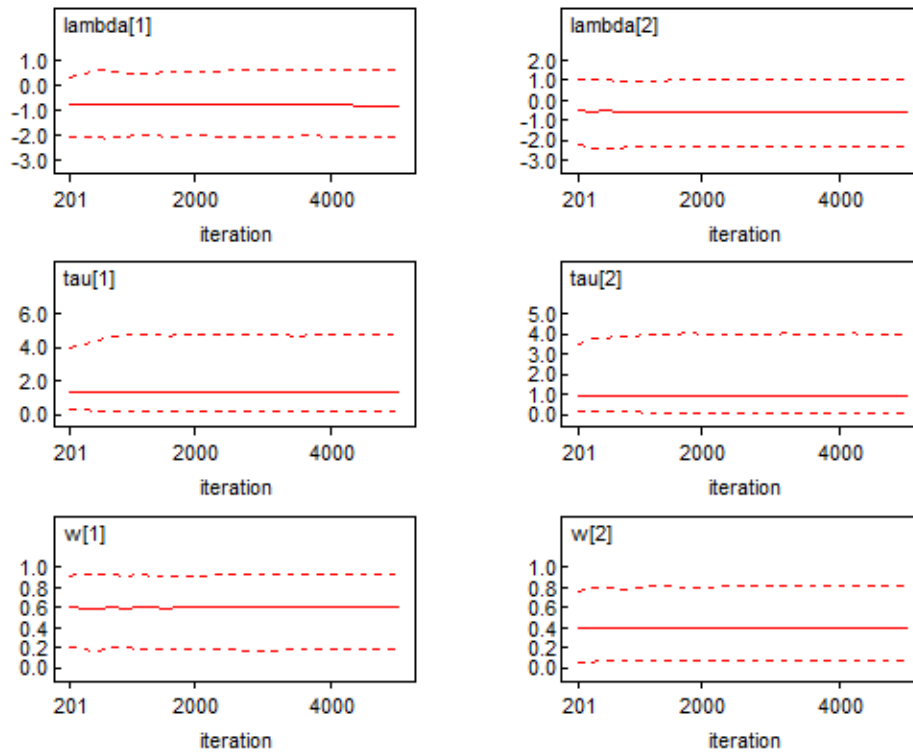
**Gambar L.5.18.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081

## Lampiran 6. (Lanjutan)



**Gambar L.5.19.** *History Data Bangkitan* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081

## Lampiran 6. (Lanjutan)



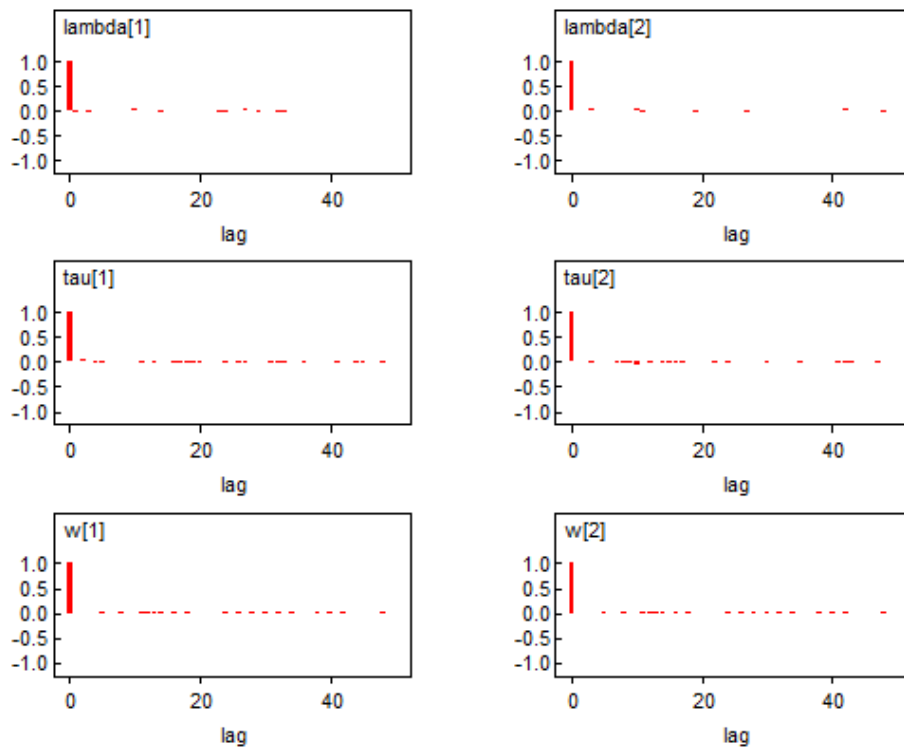
**Gambar L.5.20.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0081

## 6. ID Gen LS0752 (Pembangkitan Parameter 5.000 Kali)

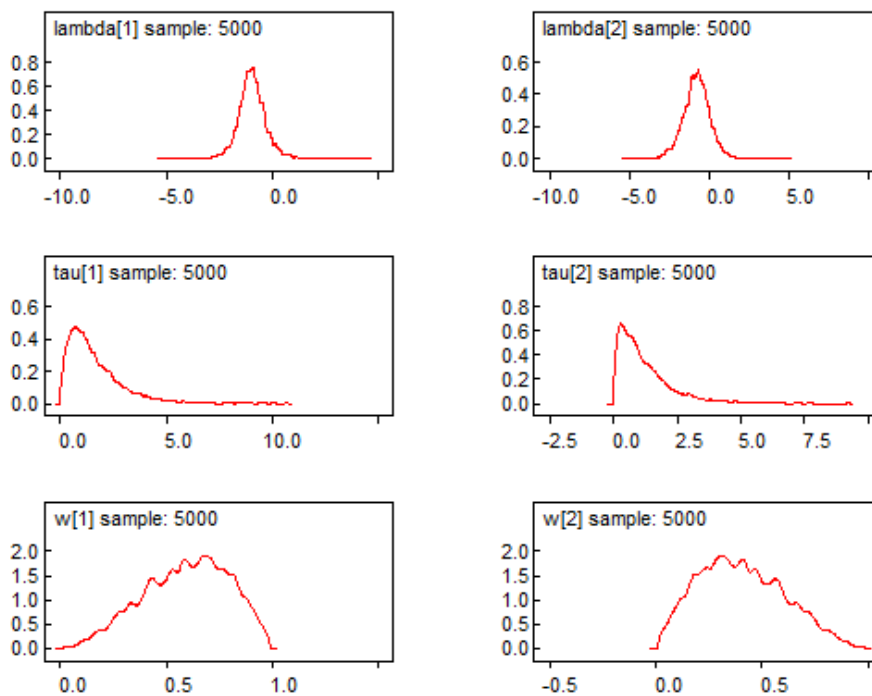
**Tabel L.5.6.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0752

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,9508	0,644	0,0079	-2,209	-0,9722	0,4038	1	5.000
lambda[2]	-0,7849	0,8312	0,0111	-2,515	-0,7625	0,8218	1	5.000
tau[1]	1,591	1,236	0,0179	0,1576	1,274	4,781	1	5.000
tau[2]	1,198	1,039	0,0134	0,0824	0,905	4,041	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

**Lampiran 6. (Lanjutan)**

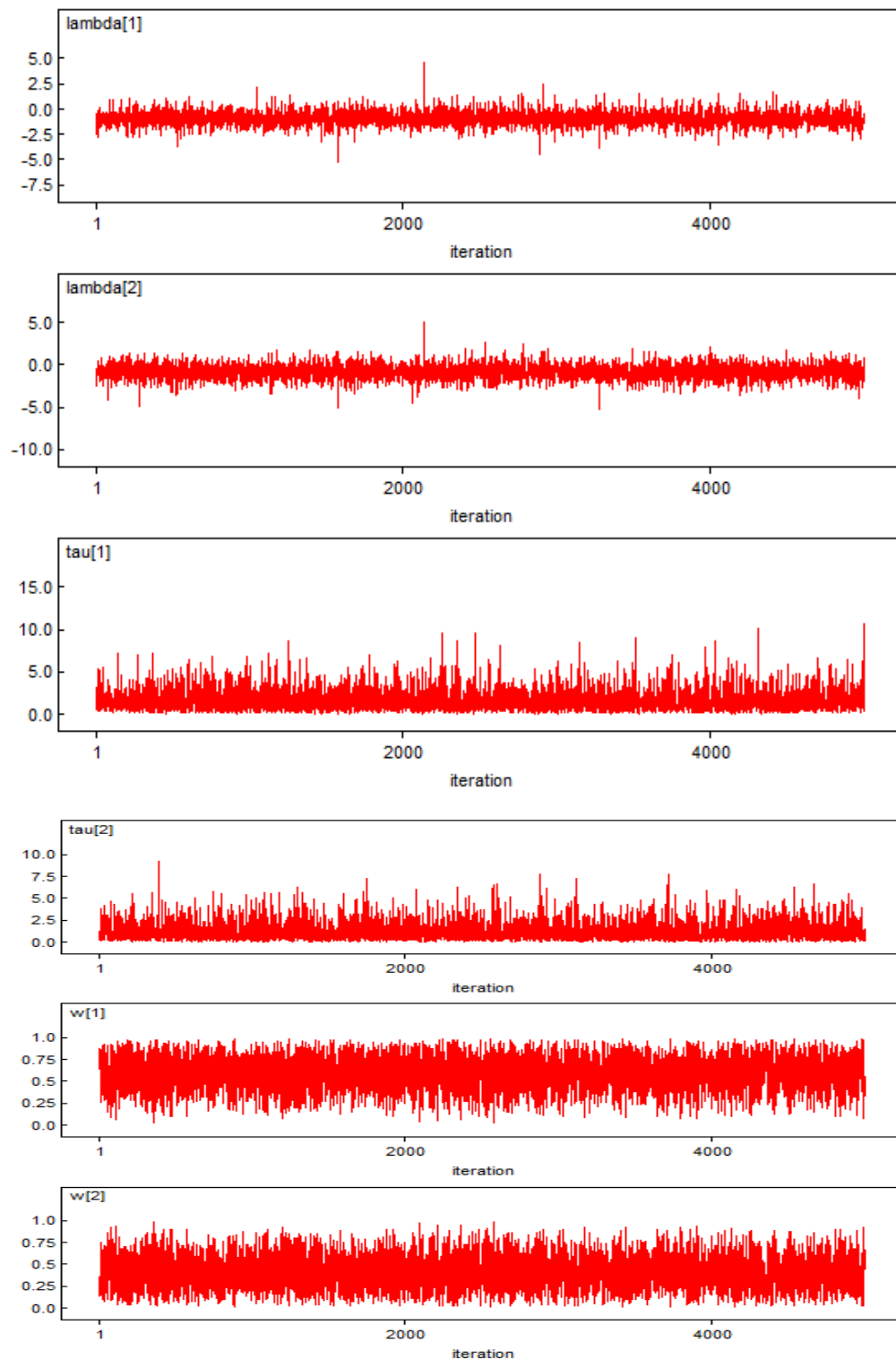


**Gambar L.5.21.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752



**Gambar L.5.22.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752

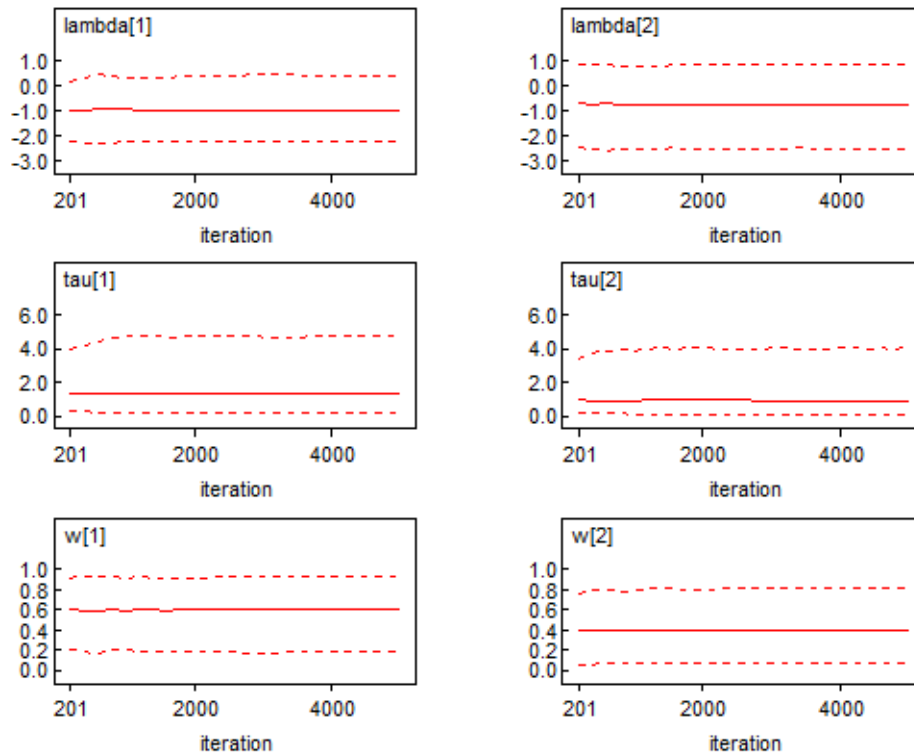
## Lampiran 6. (Lanjutan)



**Gambar L.5.23.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752*



## Lampiran 6. (Lanjutan)



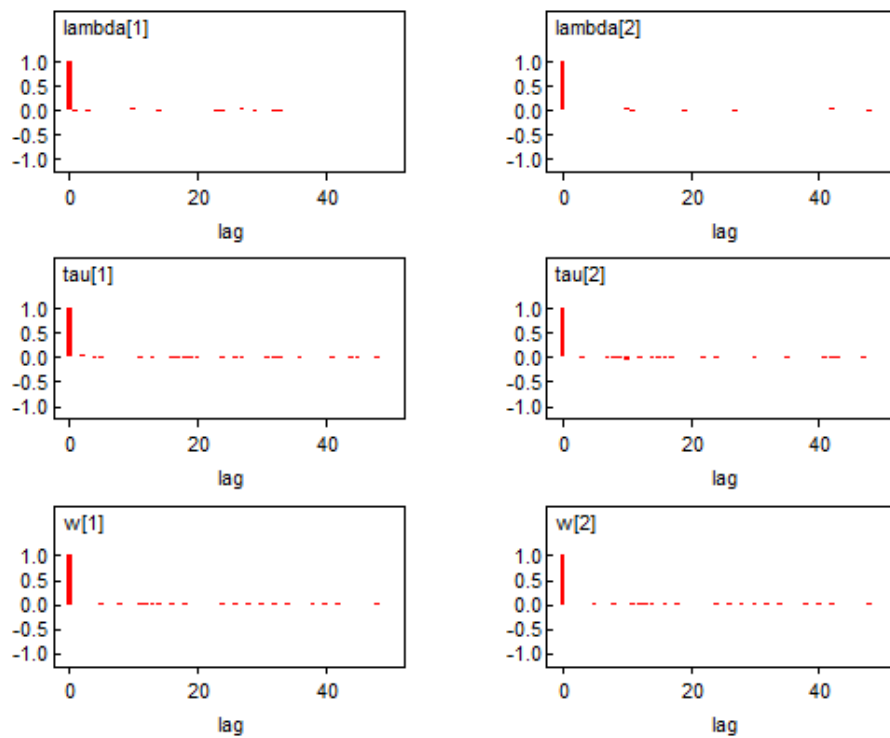
**Gambar L.5.24.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0752

## 7. ID Gen LS0688 (Pembangkitan Parameter 5.000 Kali)

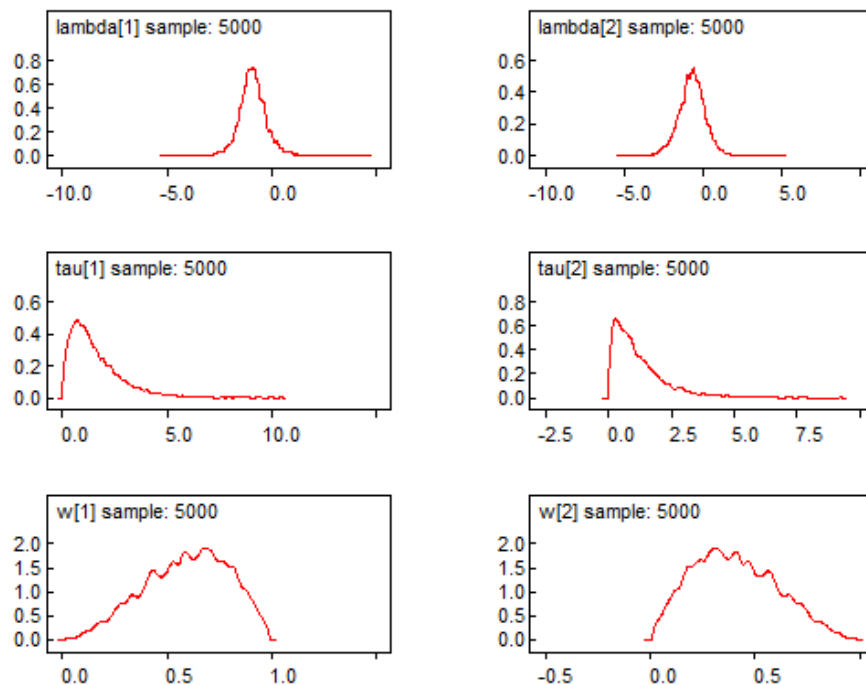
**Tabel L.5.7.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen LS0688

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,8905	0,6494	0,0079	-2,148	-0,9144	0,4981	1	5.000
lambda[2]	-0,6899	0,8364	0,0111	-2,445	-0,6642	0,9258	1	5.000
tau[1]	1,569	1,221	0,0177	0,1538	1,254	4,714	1	5.000
tau[2]	1,192	1,037	0,0134	0,0817	0,9017	4,006	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

### Lampiran 6. (Lanjutan)

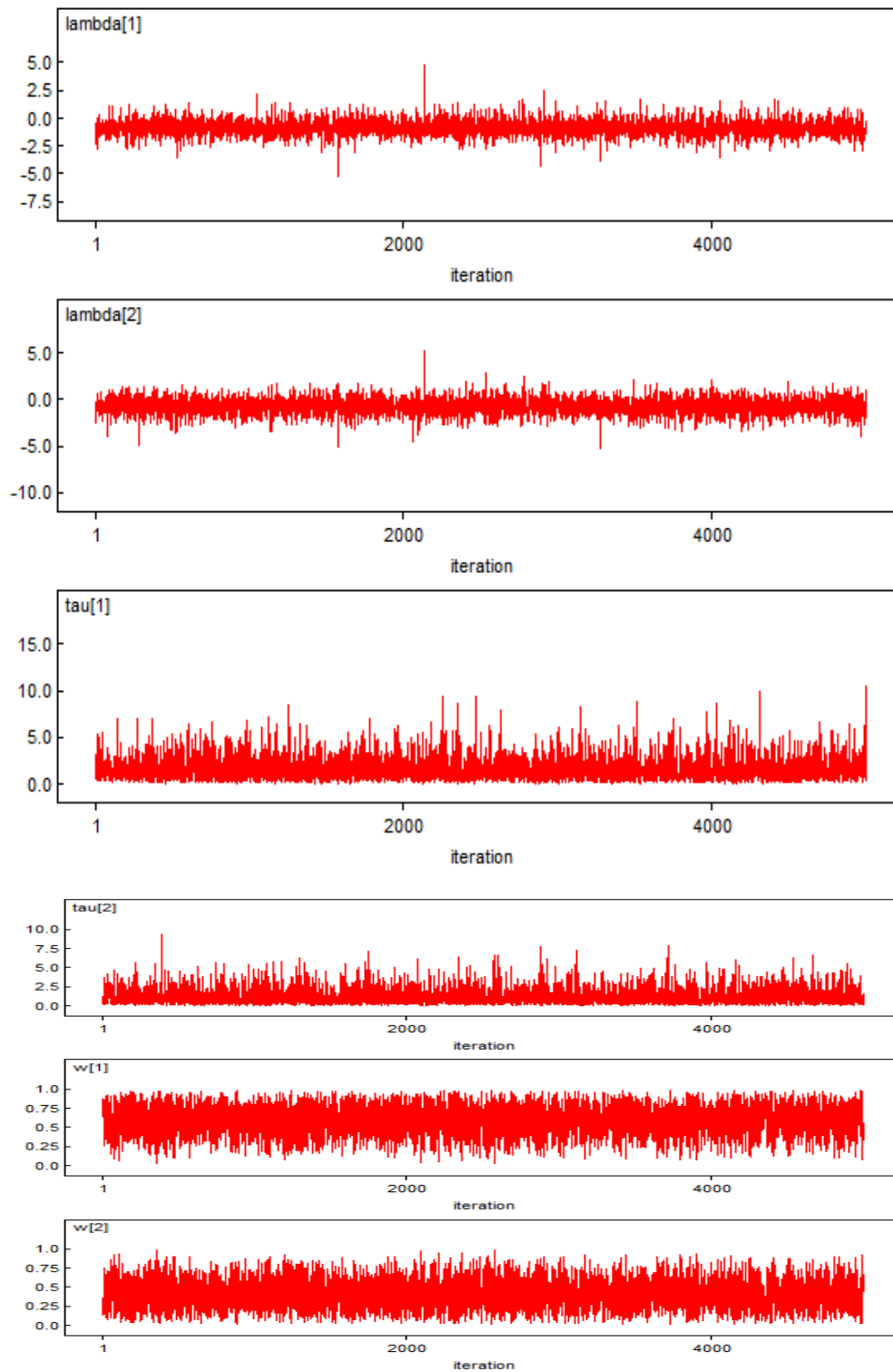


**Gambar L.5.25.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688



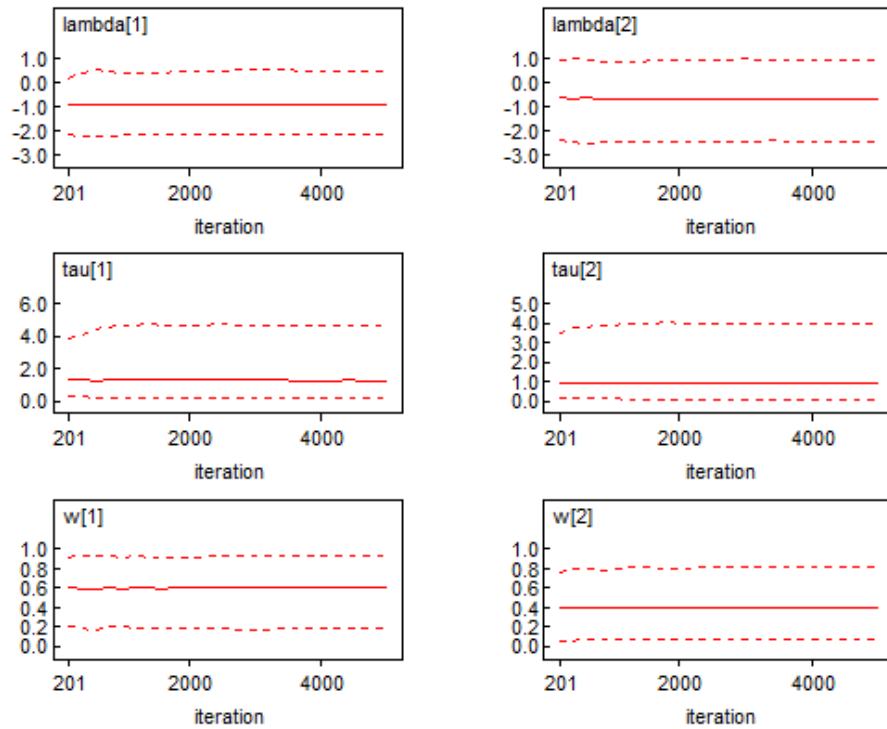
**Gambar L.5.26.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688

## Lampiran 6. (Lanjutan)



**Gambar L.5.27.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688

## Lampiran 6. (Lanjutan)



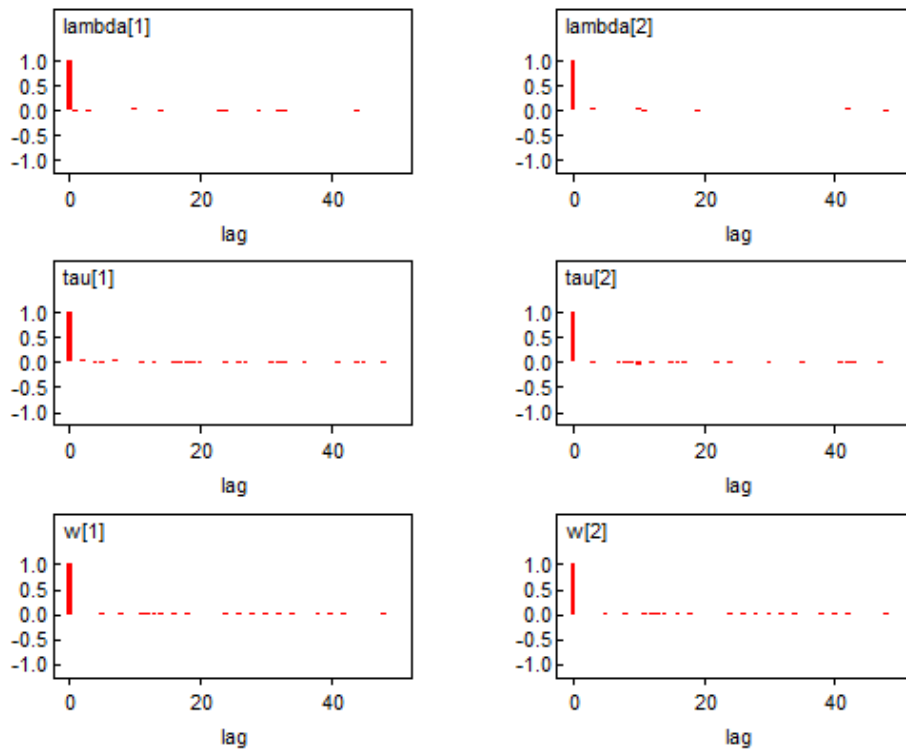
**Gambar L.5.28.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen LS0688

## 8. ID Gen U273 (Pembangkitan Parameter 5.000 Kali)

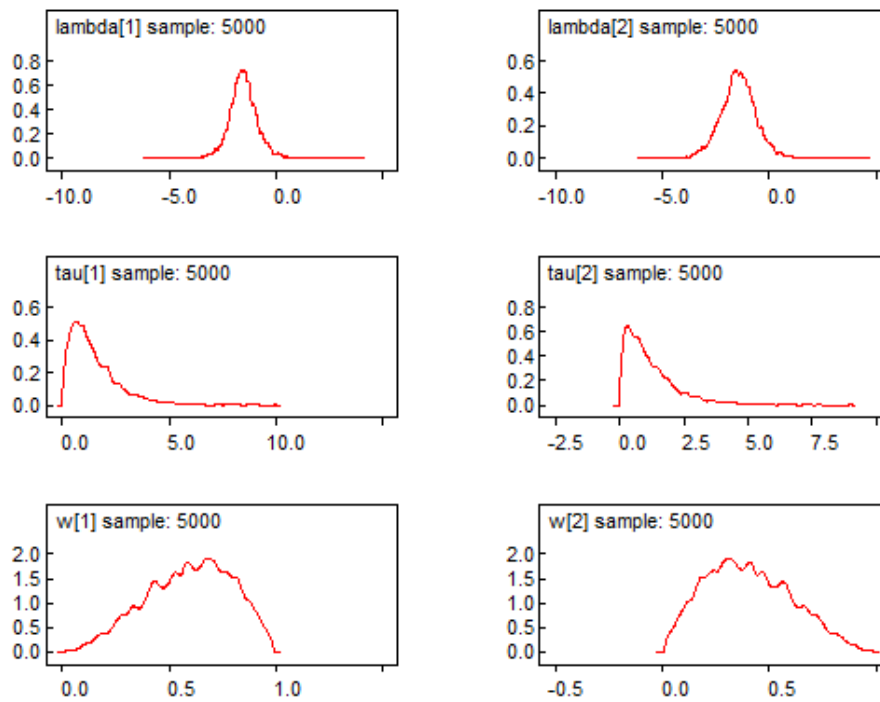
**Tabel L.5.8.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U273

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
$\lambda[1]$	-1,517	0,6593	0,0082	-2,83	-1,535	-0,1658	1	5.000
$\lambda[2]$	-1,448	0,8296	0,0111	-3,129	-1,441	0,1903	1	5.000
$\tau[1]$	1,463	1,13	0,0162	0,1456	1,169	4,394	1	5.000
$\tau[2]$	1,207	1,041	0,0135	0,0822	0,9137	4,053	1	5.000
$w[1]$	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
$w[2]$	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

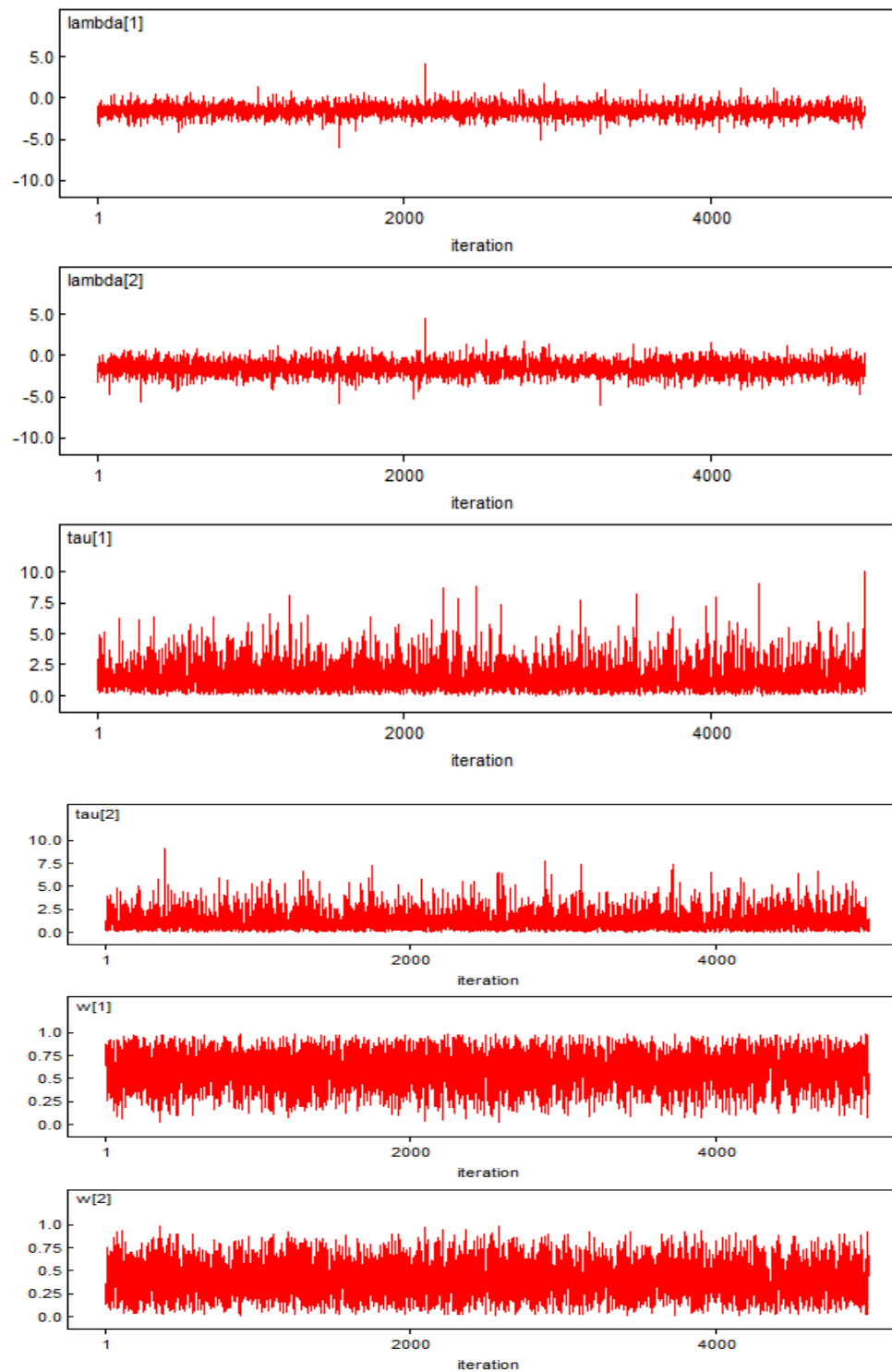


**Gambar L.5.29.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U273



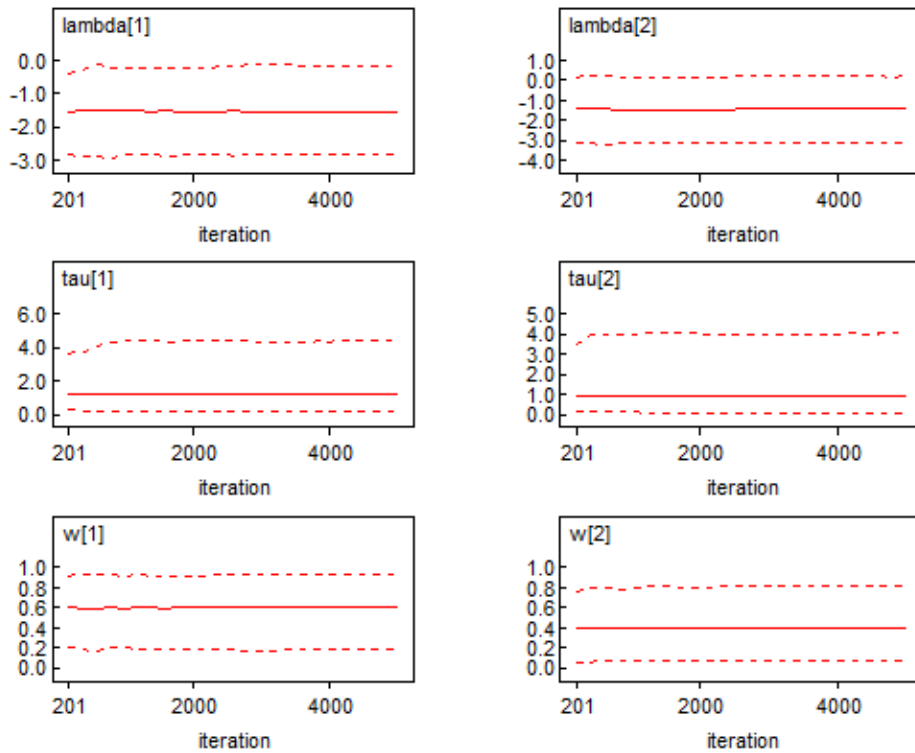
**Gambar L.5.30.** Densitas Kernel setiap Parameter BMM-Normal DuaKomponen untuk ID Gen U273

## Lampiran 6. (Lanjutan)



**Gambar L.5.31.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U273*

## Lampiran 6. (Lanjutan)



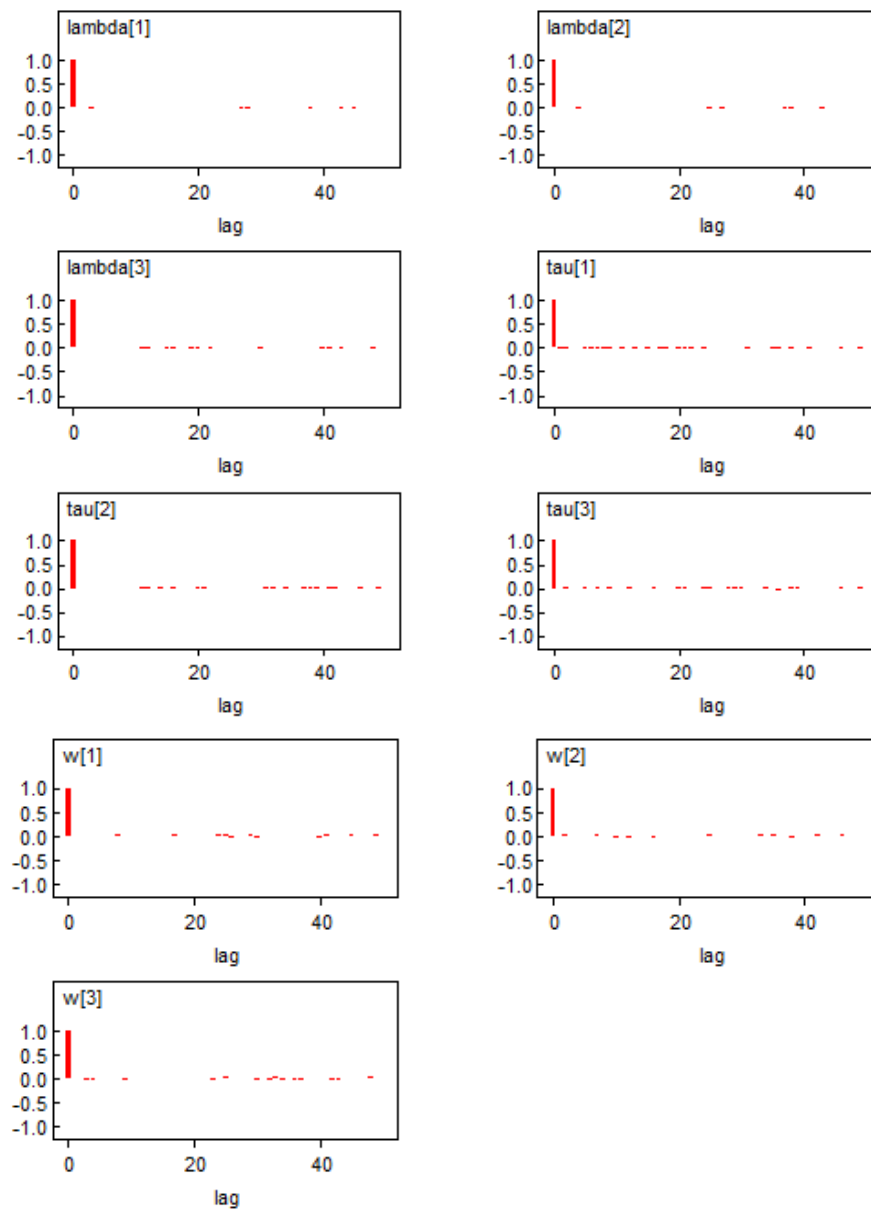
**Gambar L.5.32.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U273

## 9. ID Gen U070 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.5.9.** Estimasi Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U070

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,106	0,2002	0,0025	-1,508	-1,104	-0,7109	1	5.000
lambda[2]	-0,9943	0,1968	0,0025	-1,383	-0,9941	-0,6053	1	5.000
lambda[3]	-0,1982	0,2498	0,0038	-0,682	-0,2017	0,2888	1	5.000
tau[1]	11,96	0,8308	0,011	10,32	11,97	13,59	1	5.000
tau[2]	14,92	0,8169	0,0118	13,33	14,91	16,55	1	5.000
tau[3]	15	0,8366	0,0123	13,35	15,02	16,64	1	5.000
w[1]	0,334	0,1797	0,0027	0,0525	0,3134	0,7215	1	5.000
w[2]	0,3357	0,1793	0,0027	0,0588	0,3124	0,735	1	5.000
w[3]	0,3304	0,1794	0,0027	0,0519	0,3125	0,7121	1	5.000

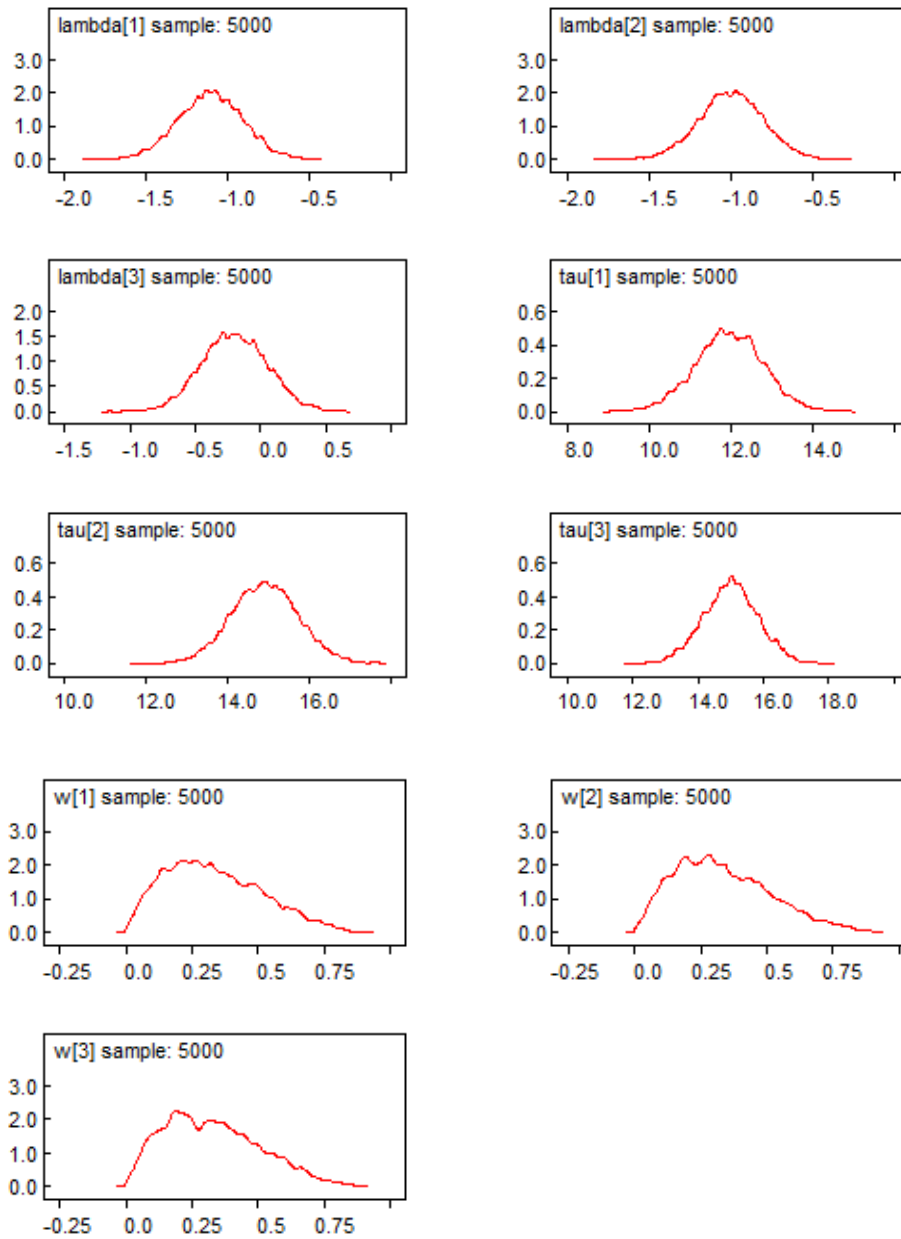
## Lampiran 6. (Lanjutan)



**Gambar L.5.33.** Fungsi Autokorelasi setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070

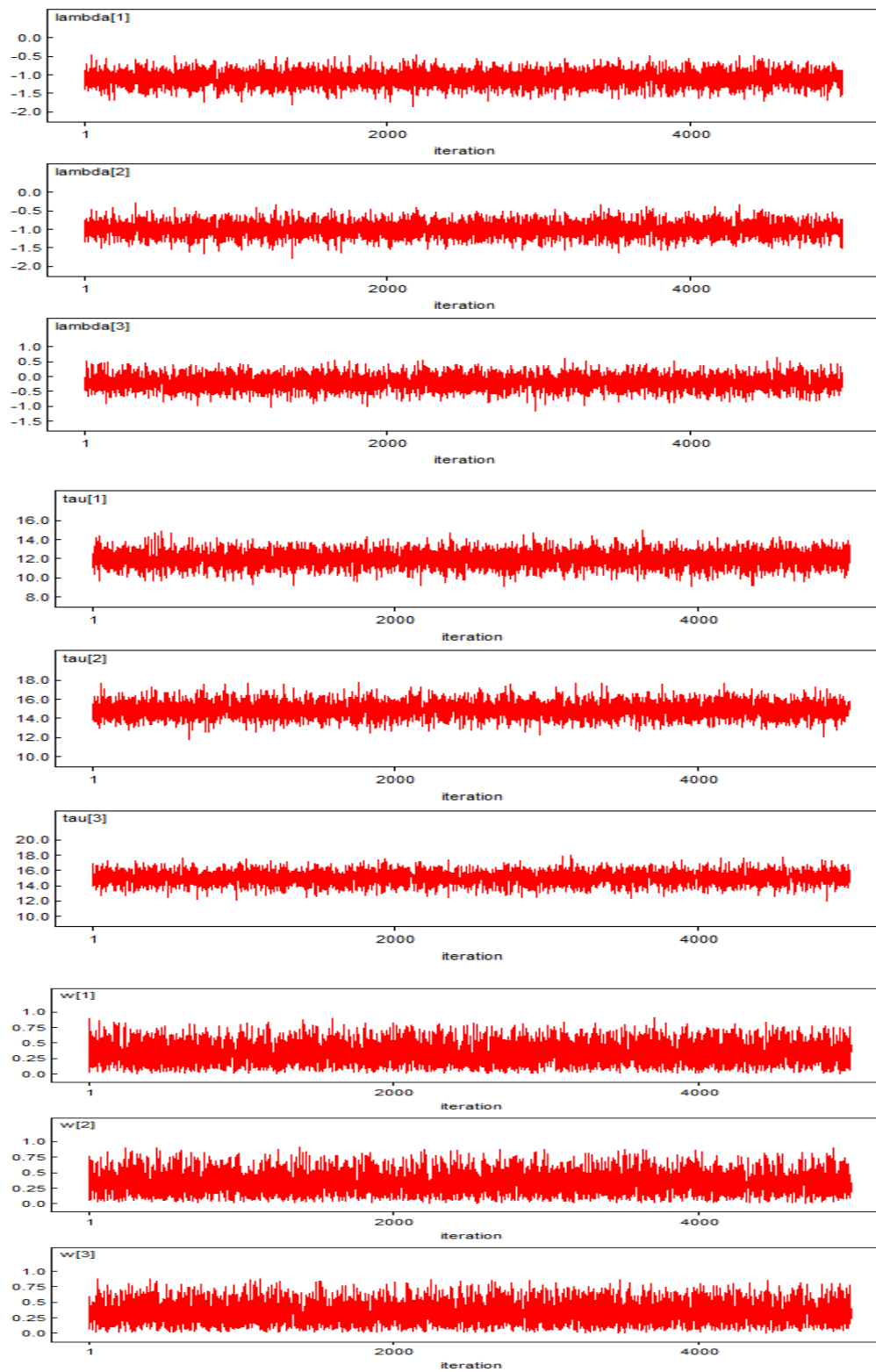


## Lampiran 6. (Lanjutan)



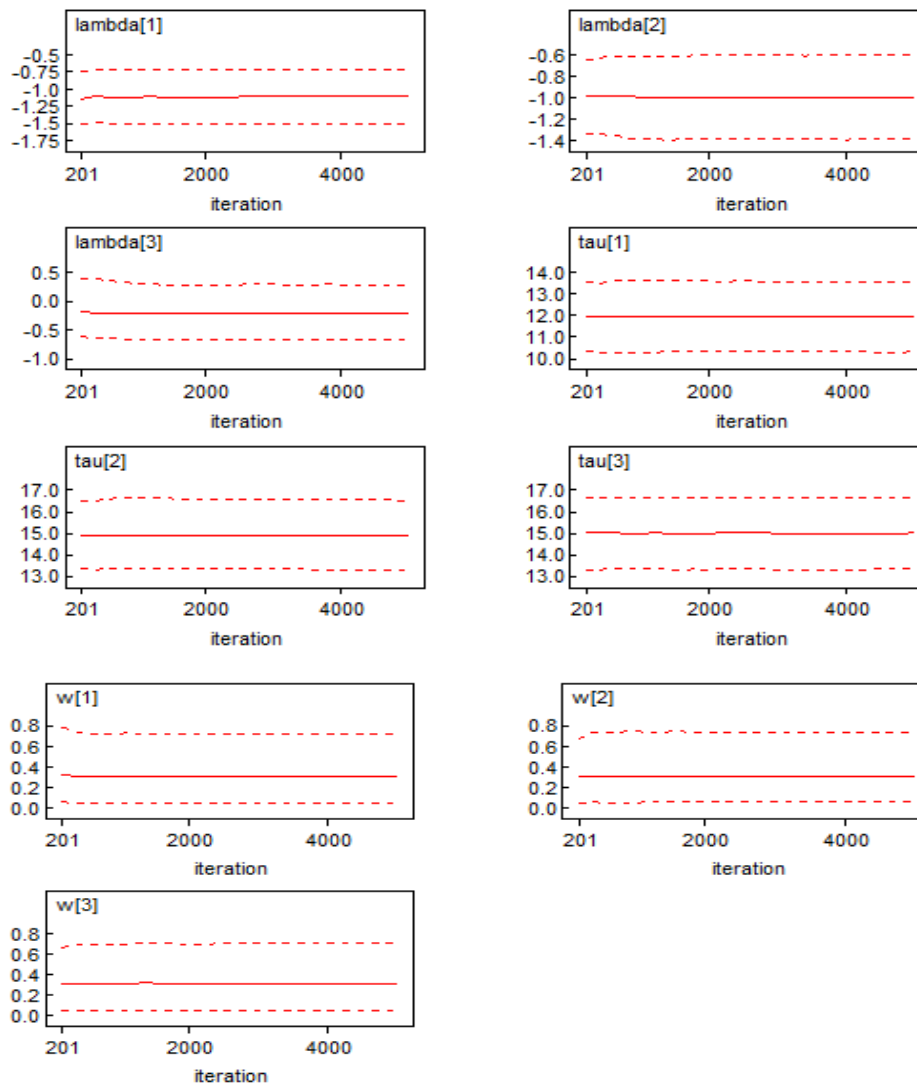
**Gambar L.5.34.** Densitas Kernel setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070

## Lampiran 6. (Lanjutan)



**Gambar L.5.35.** *History Data* Bangkitan setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070

## Lampiran 6. (Lanjutan)



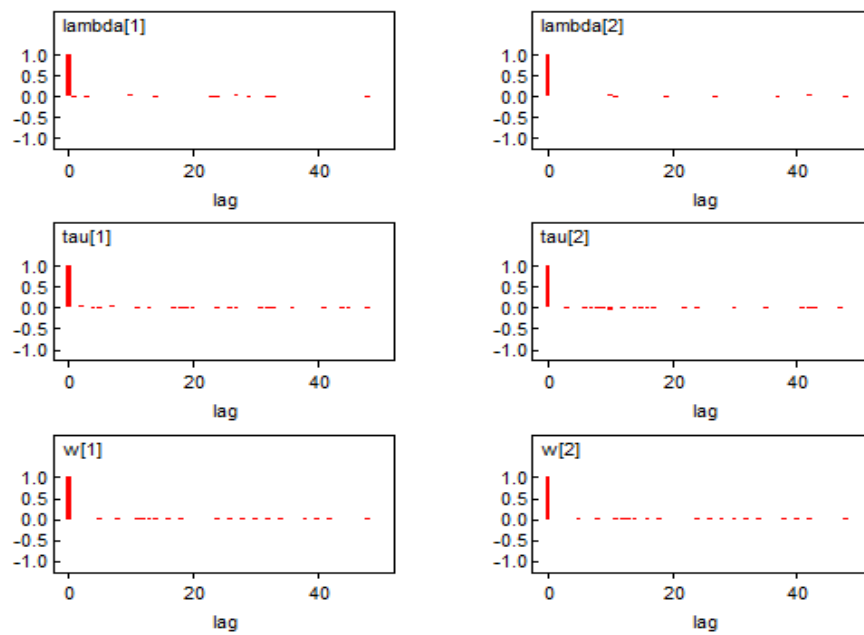
**Gambar L.5.36.** *Quantiles* setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U070

### 10. ID Gen U325 (Pembangkitan Parameter 5.000 Kali)

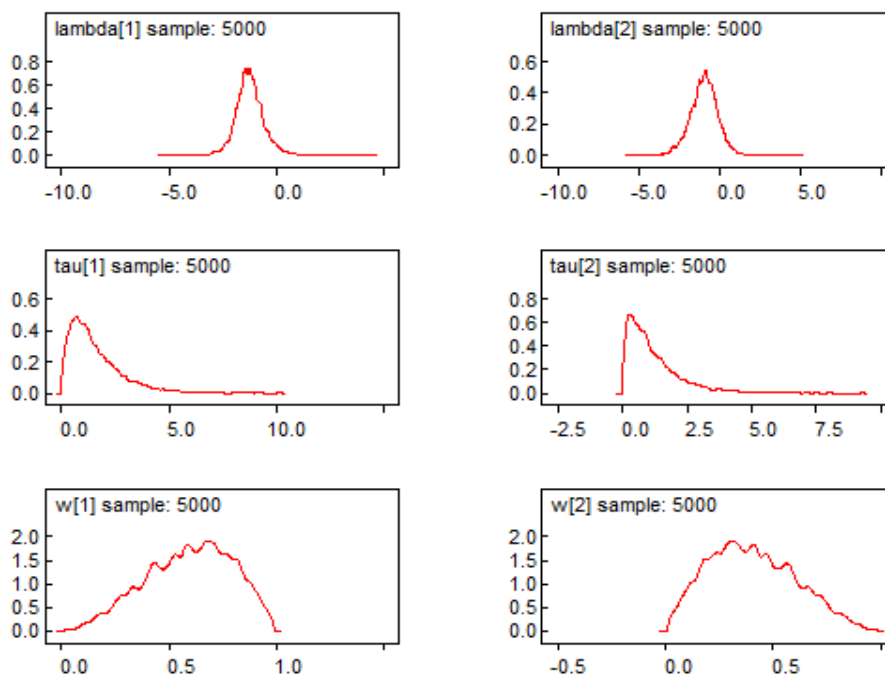
**Tabel L.5.10.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U325

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,233	0,6595	0,0079	-2,461	-1,266	0,2009	1	5.000
lambda[2]	-0,955	0,8499	0,0113	-2,759	-0,9213	0,6616	1	5.000
tau[1]	1,563	1,223	0,0179	0,1476	1,244	4,753	1	5.000
tau[2]	1,173	1,031	0,0133	0,0809	0,8816	3,966	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

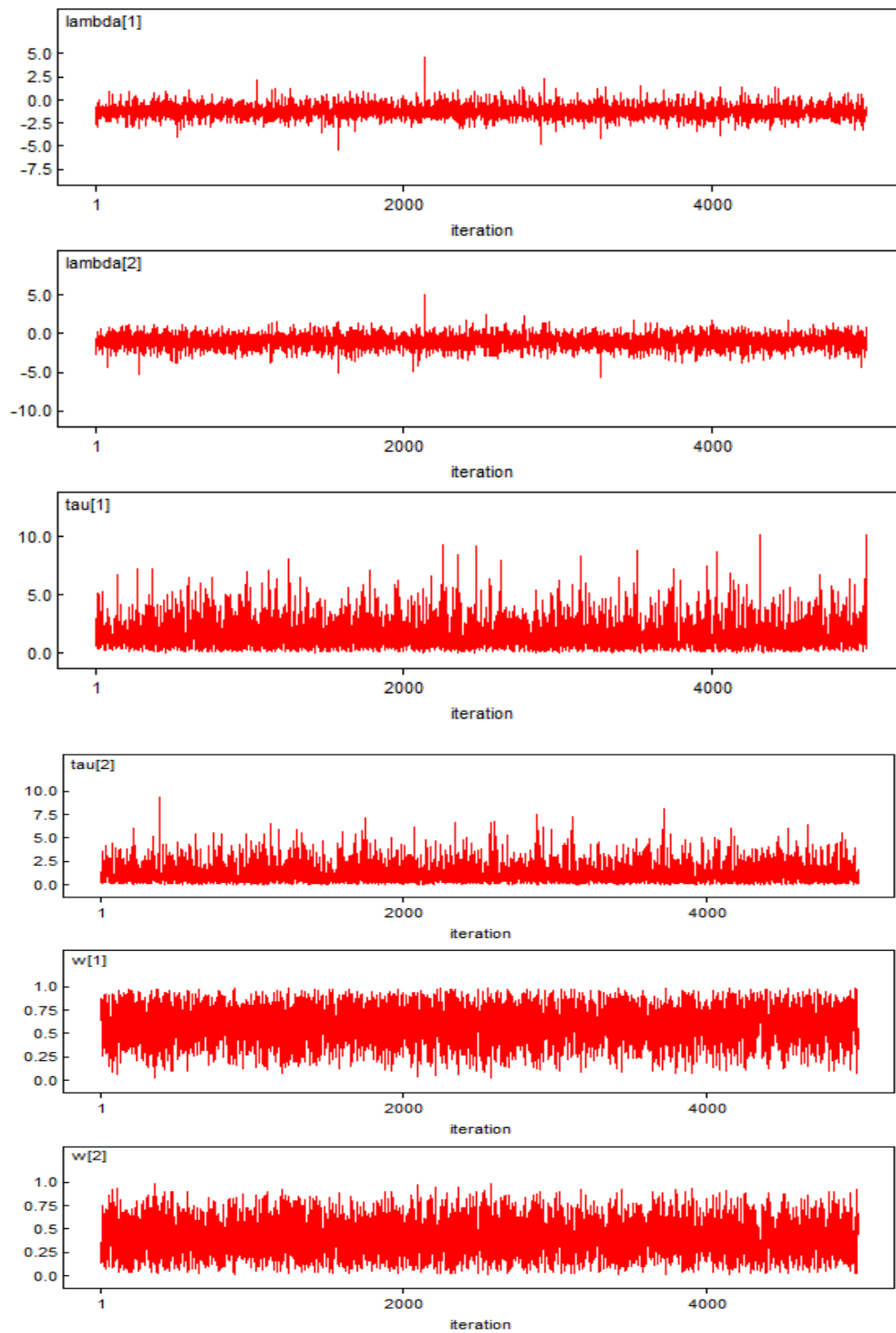


**Gambar L.5.37.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325



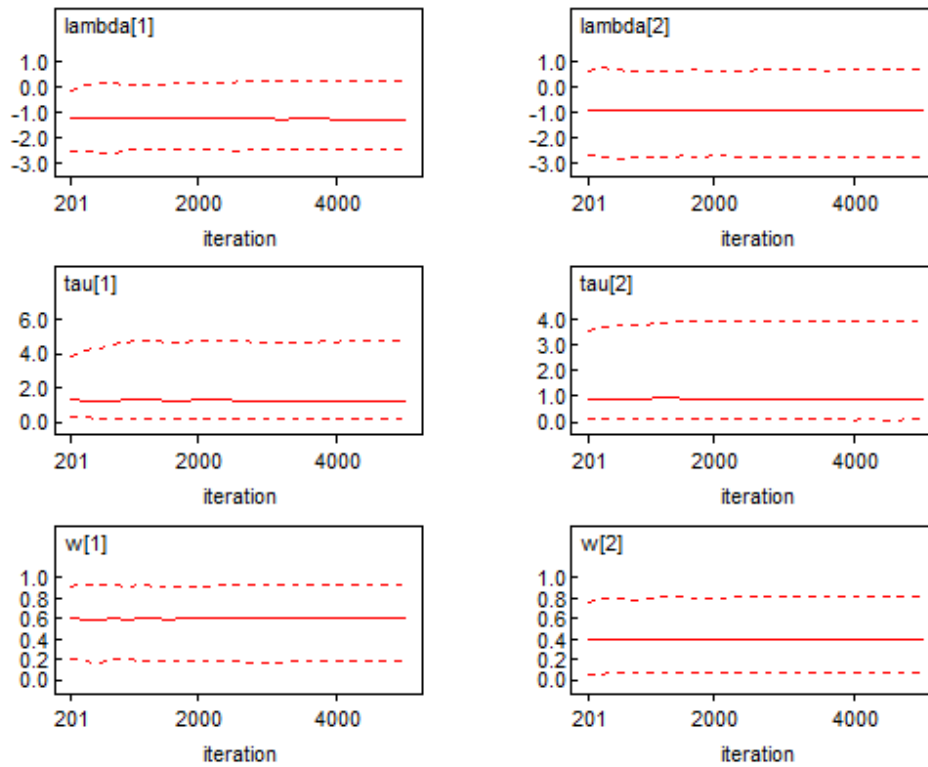
**Gambar L.5.38.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325

## Lampiran 6. (Lanjutan)



**Gambar L.5.39.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325

## Lampiran 6. (Lanjutan)



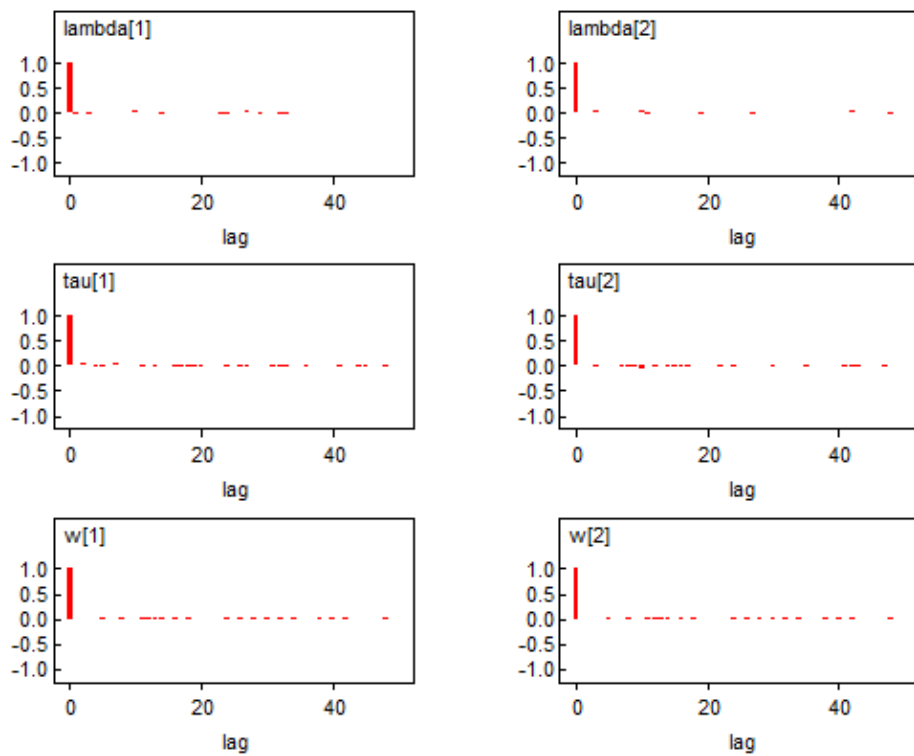
**Gambar L.5.40.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U325

### 11. ID Gen U385 (Pembangkitan Parameter 5.000 Kali)

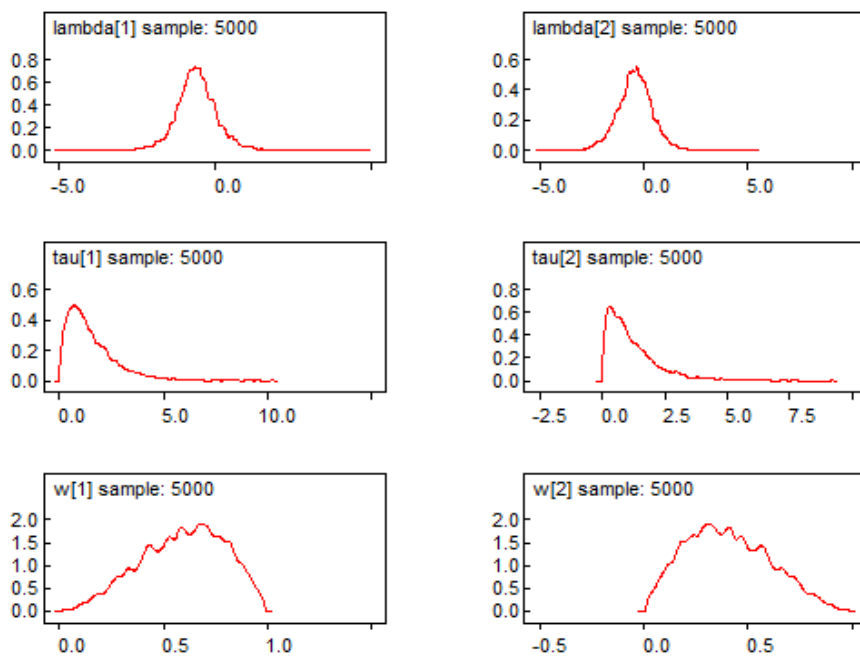
**Tabel L.5.11.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U385

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5936	0,653	0,008	-1,883	-0,6166	0,7671	1	5.000
lambda[2]	-0,4394	0,8329	0,0111	-2,168	-0,4206	1,179	1	5.000
tau[1]	1,516	1,175	0,017	0,1505	1,215	4,554	1	5.000
tau[2]	1,199	1,039	0,0134	0,0826	0,9059	4,041	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

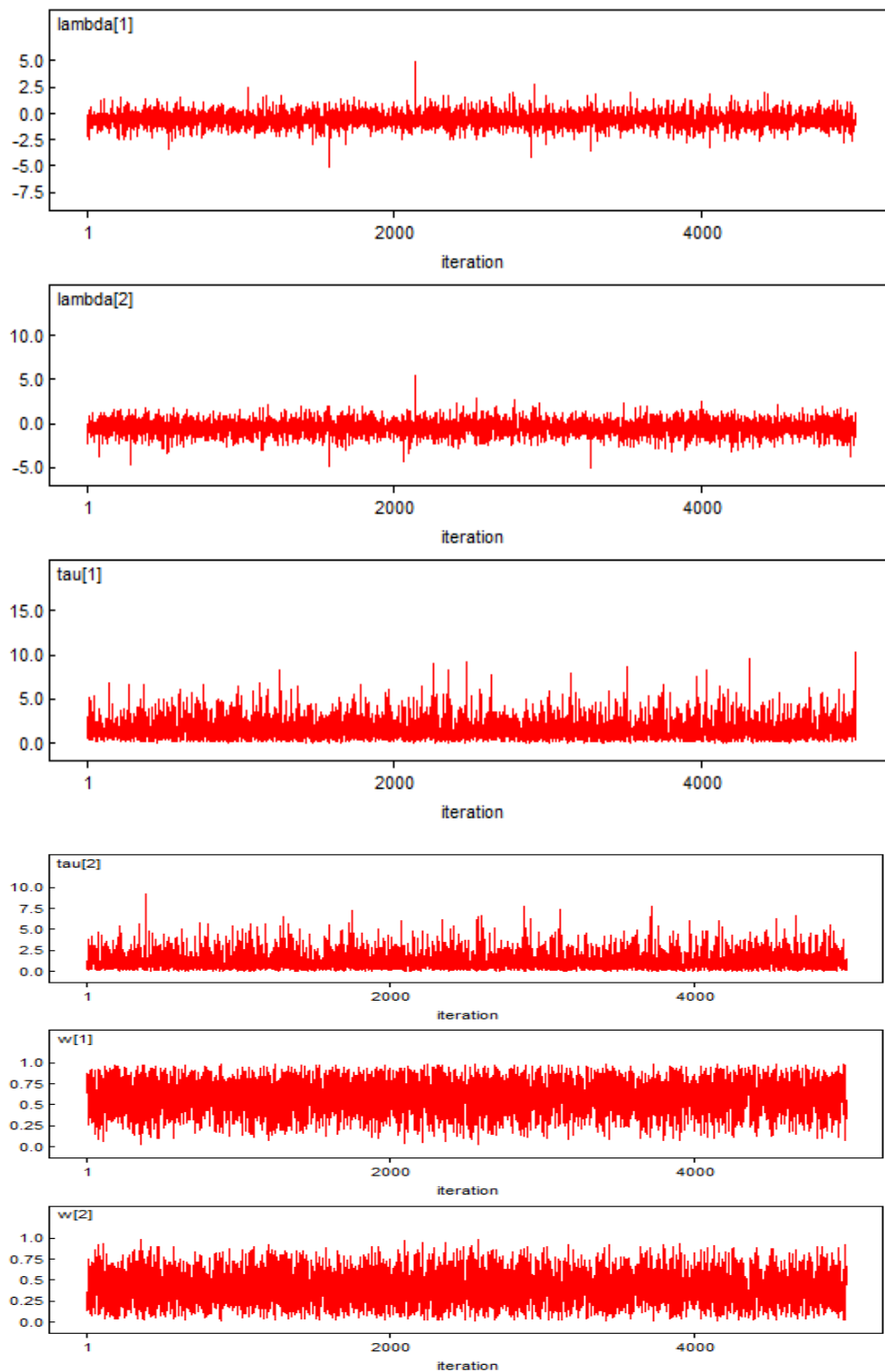


**Gambar L.5.41.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385



**Gambar L.5.42.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385

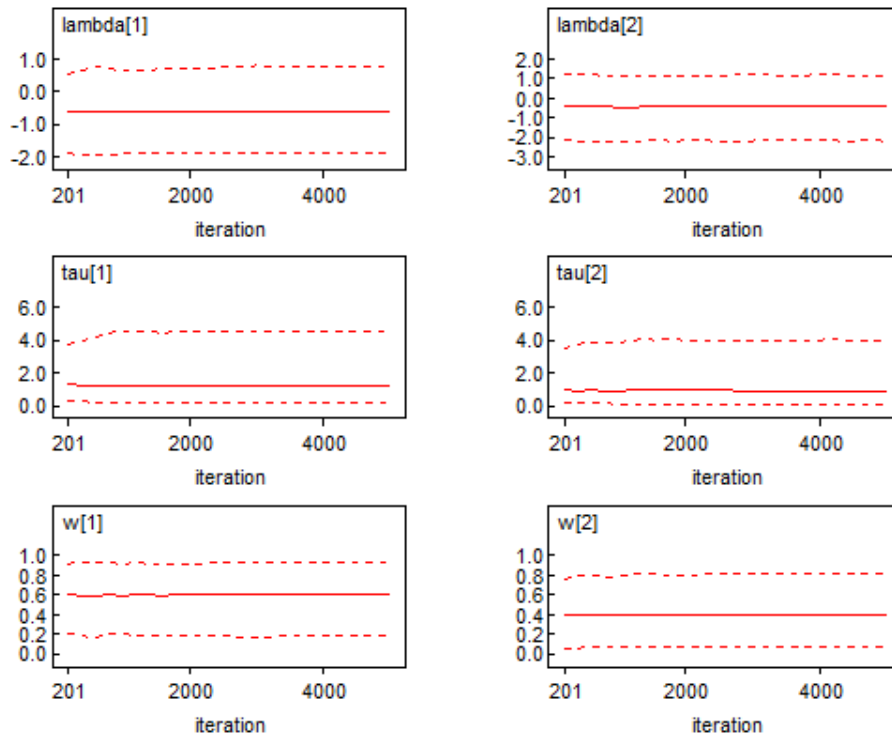
## Lampiran 6. (Lanjutan)



**Gambar L.5.43.** *History Data* Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385



## Lampiran 6. (Lanjutan)



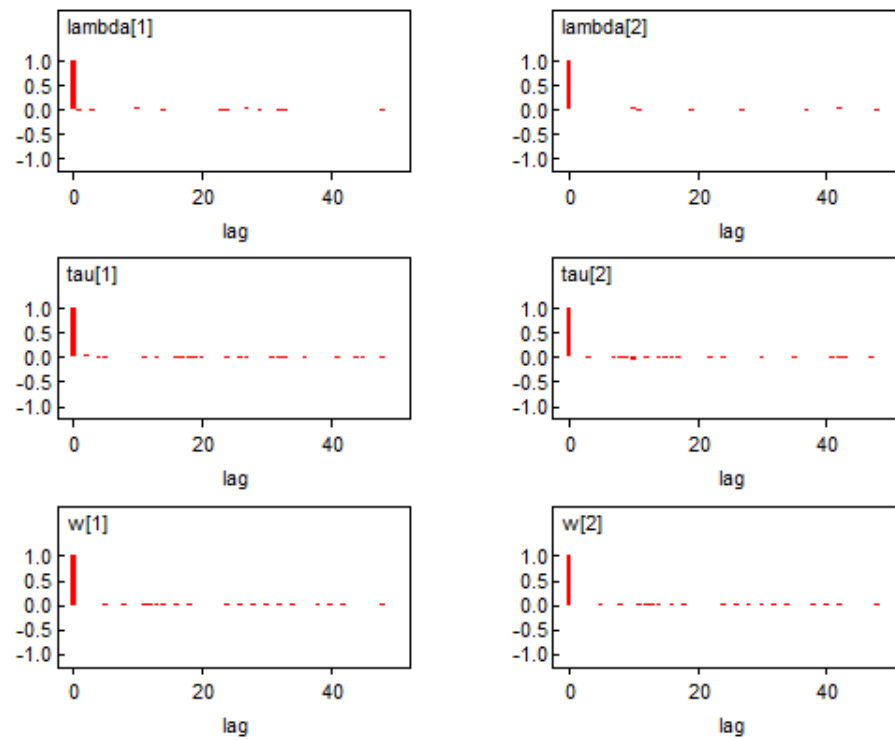
**Gambar L.5.44.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U385

## 12. ID Gen U447 (Pembangkitan Parameter 5.000 Kali)

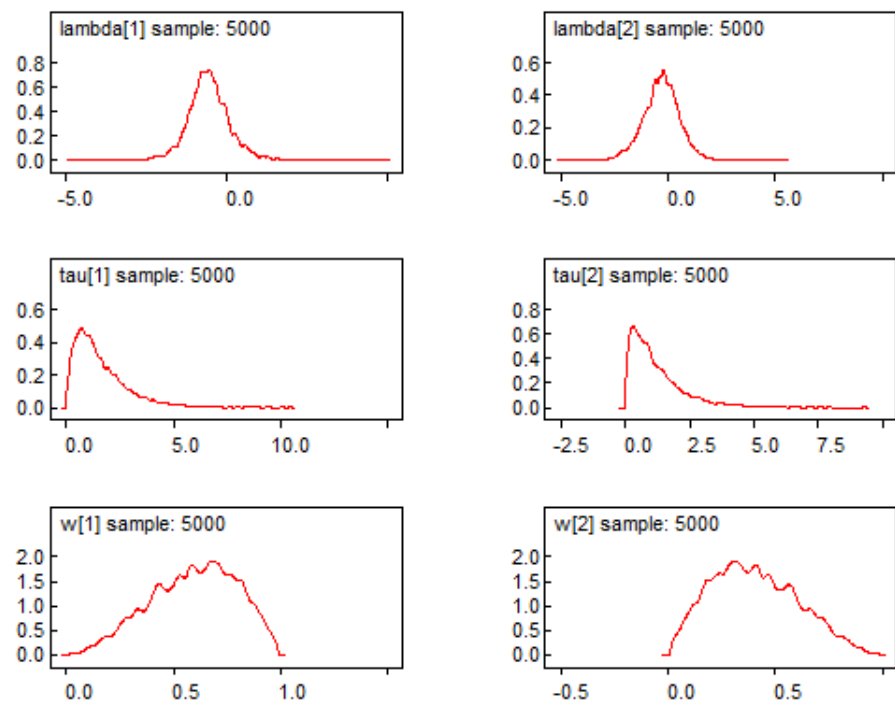
**Tabel L.5.12.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen U447

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
$\lambda[1]$	-0,5582	0,6498	0,0079	-1,809	-0,5831	0,8357	1	5.000
$\lambda[2]$	-0,3231	0,8407	0,0112	-2,097	-0,2929	1,292	1	5.000
$\tau[1]$	1,579	1,23	0,0179	0,1537	1,259	4,746	1	5.000
$\tau[2]$	1,185	1,035	0,0134	0,0809	0,892	4,03	1	5.000
$w[1]$	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
$w[2]$	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

### Lampiran 6. (Lanjutan)

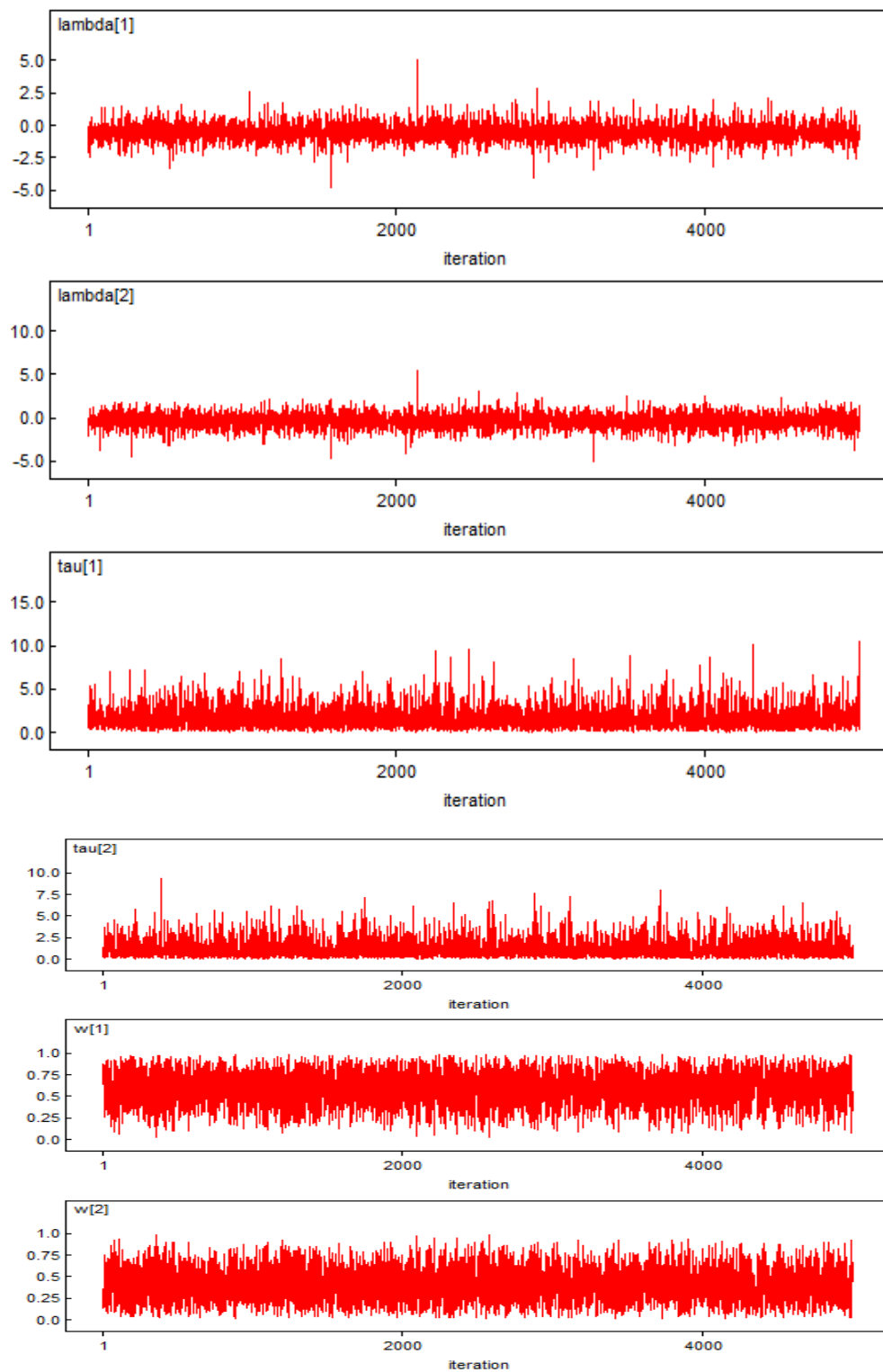


**Gambar L.5.44.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447



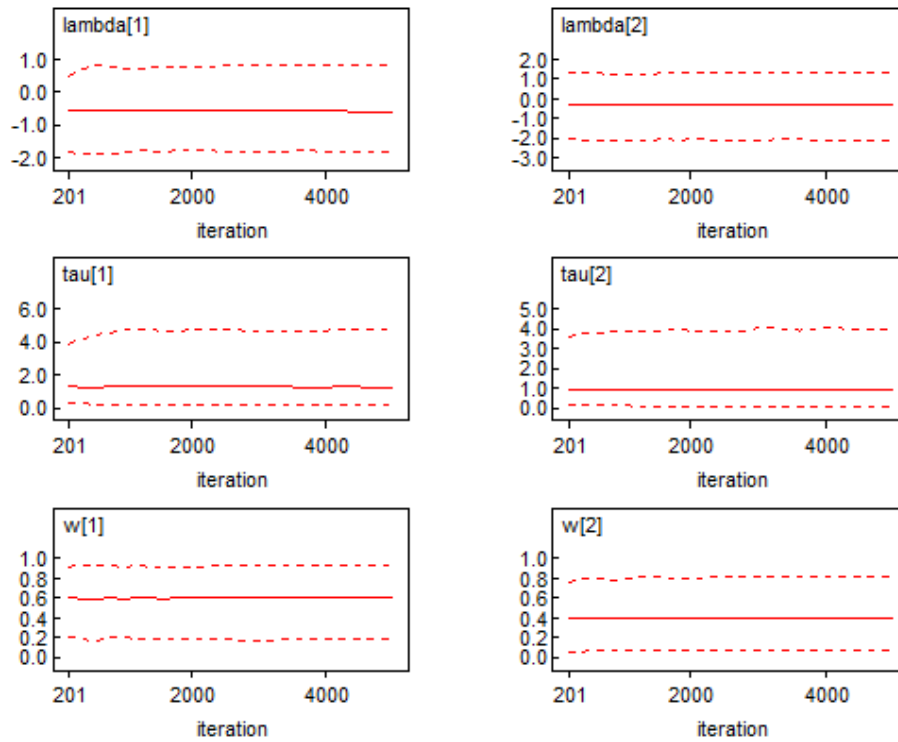
**Gambar L.5.45.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447

## Lampiran 6. (Lanjutan)



**Gambar L.5.46.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447*

## Lampiran 6. (Lanjutan)



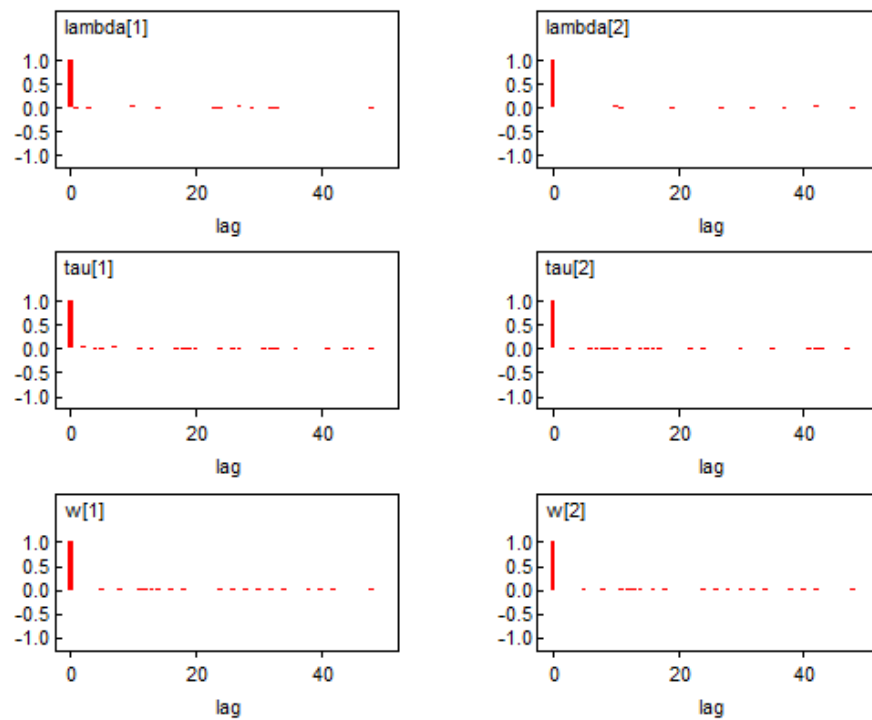
**Gambar L.5.47.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen U447

### 13. ID Gen CA0426 (Pembangkitan Parameter 5.000 Kali)

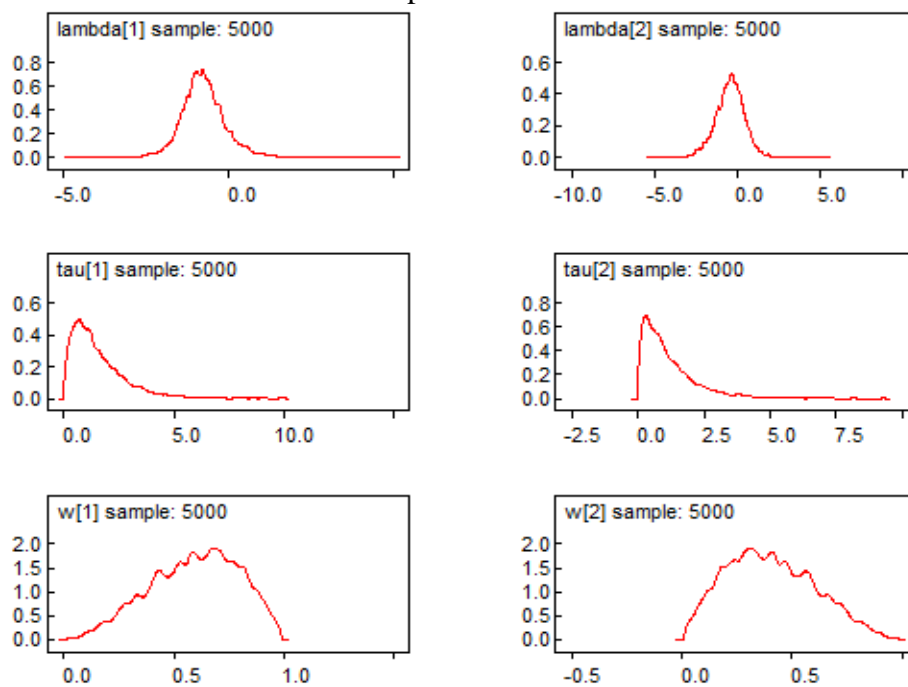
**Tabel L.5.13.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA0426

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,7652	0,6685	0,008	-1,998	-0,8031	0,696	1	5.000
lambda[2]	-0,4285	0,8633	0,0114	-2,269	-0,3885	1,196	1	5.000
tau[1]	1,538	1,207	0,0177	0,1452	1,223	4,702	1	5.000
tau[2]	1,155	1,024	0,0133	0,0791	0,8622	3,944	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

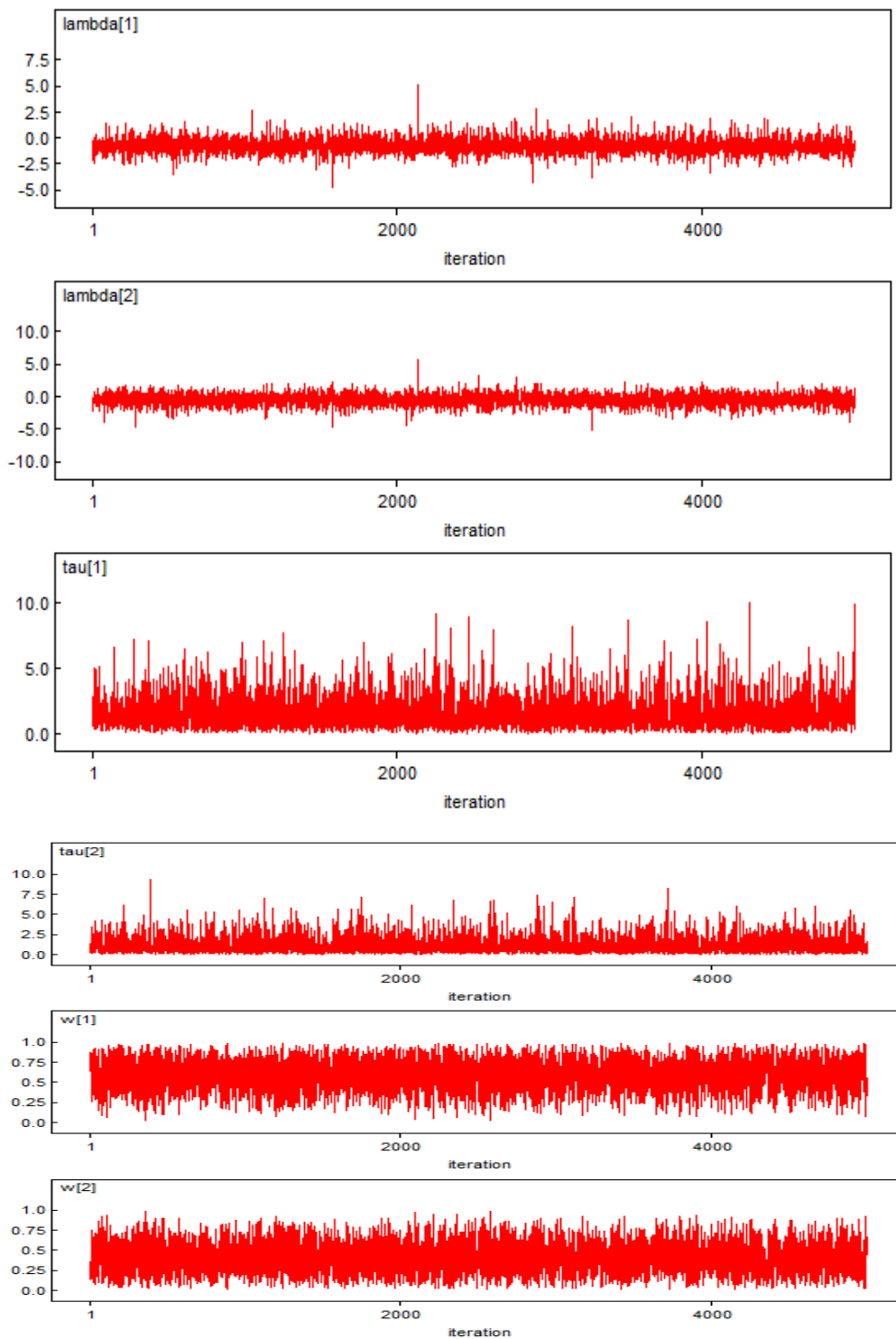


**Gambar L.5.49.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426



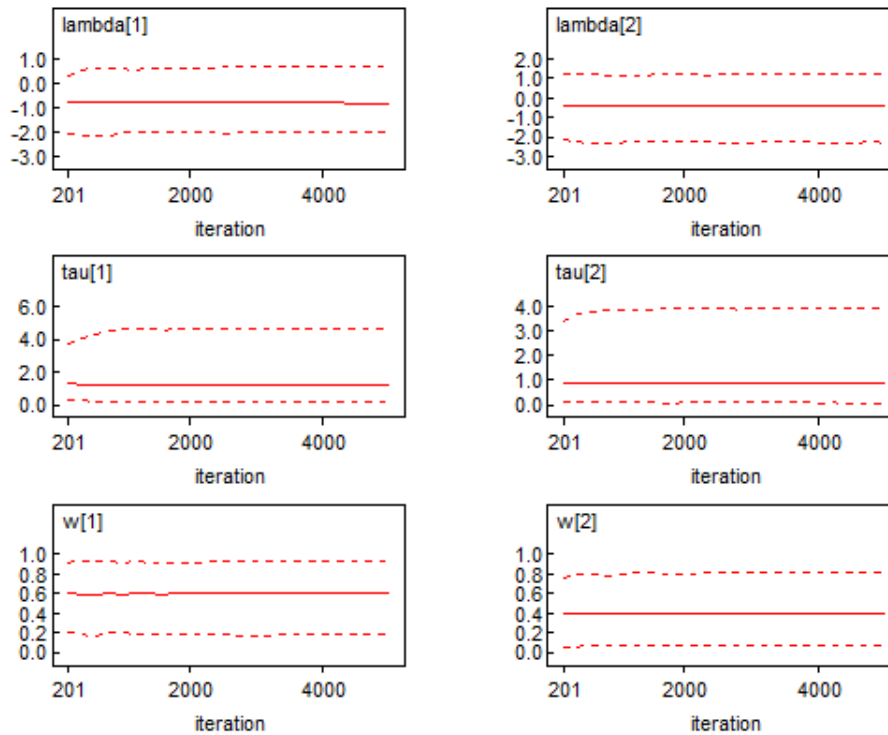
**Gambar L.5.51.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426

## Lampiran 6. (Lanjutan)



**Gambar L.5.52.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426*

## Lampiran 6. (Lanjutan)



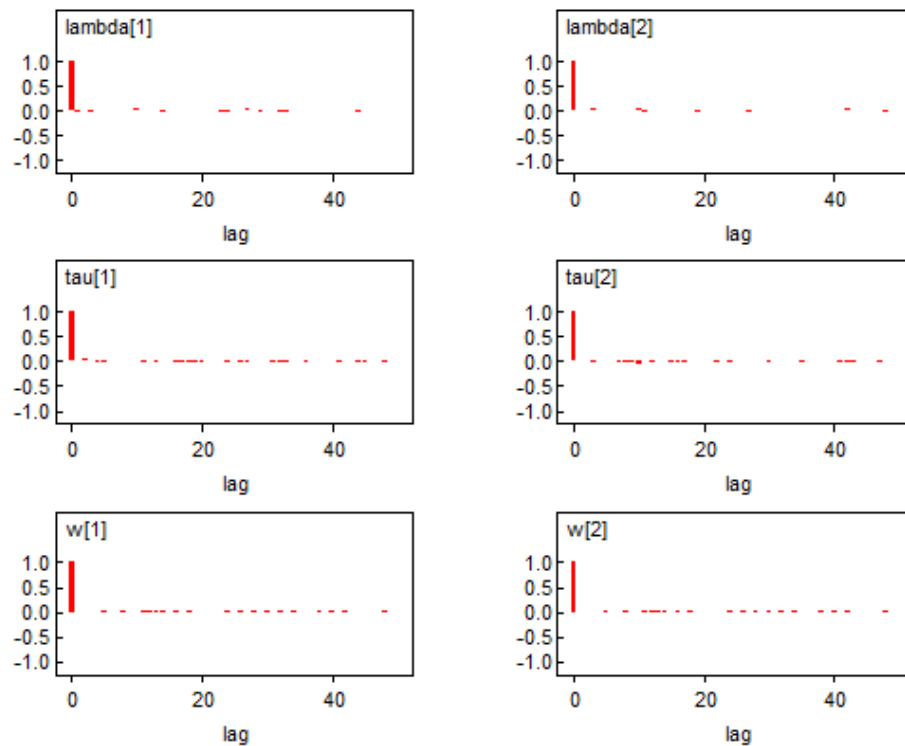
**Gambar L.5.53.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0426

### 14. ID Gen CA0534 (Pembangkitan Parameter 5.000 Kali)

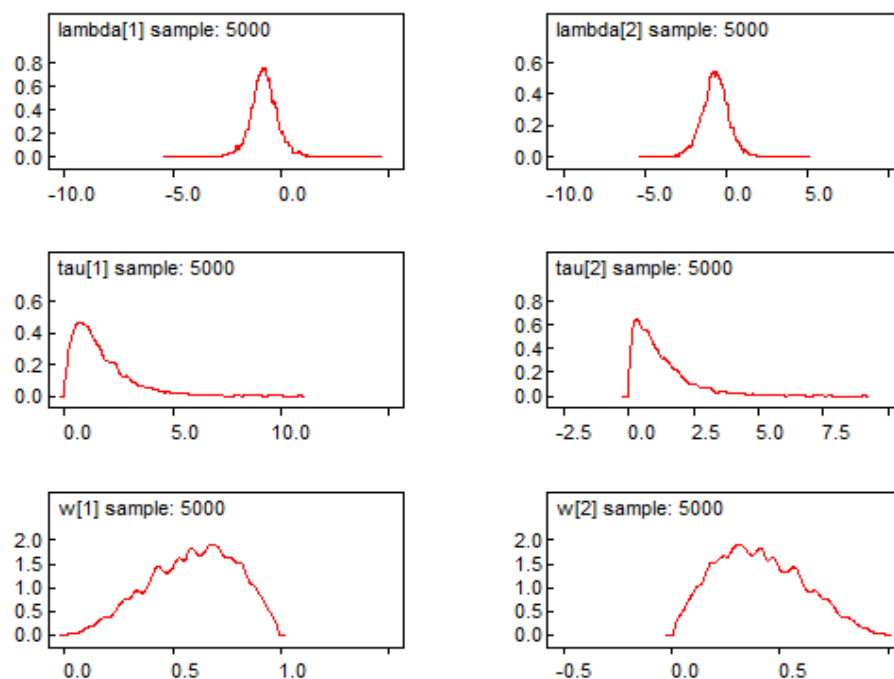
**Tabel L.5.14.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA0534

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
$\lambda[1]$	-0,7609	0,639	0,0079	-2,035	-0,7775	0,5431	1	5.000
$\lambda[2]$	-0,6678	0,8245	0,0111	-2,359	-0,6569	0,9495	1	5.000
$\tau[1]$	1,598	1,237	0,0177	0,1572	1,277	4,804	1	5.000
$\tau[2]$	1,207	1,042	0,0135	0,0833	0,9171	4,073	1	5.000
$w[1]$	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
$w[2]$	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

### Lampiran 6. (Lanjutan)



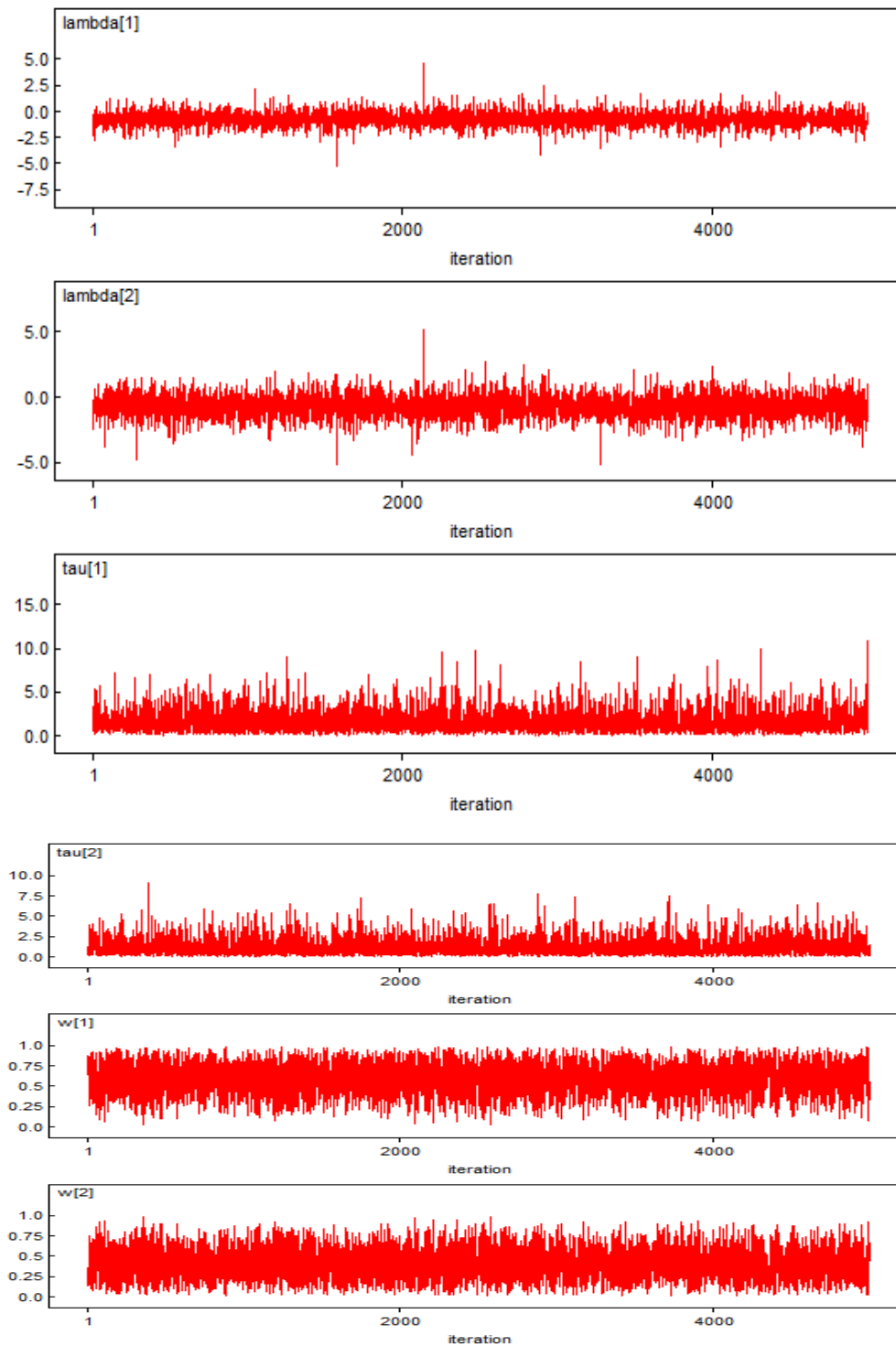
**Gambar L.5.54.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534



**Gambar L.5.55.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534

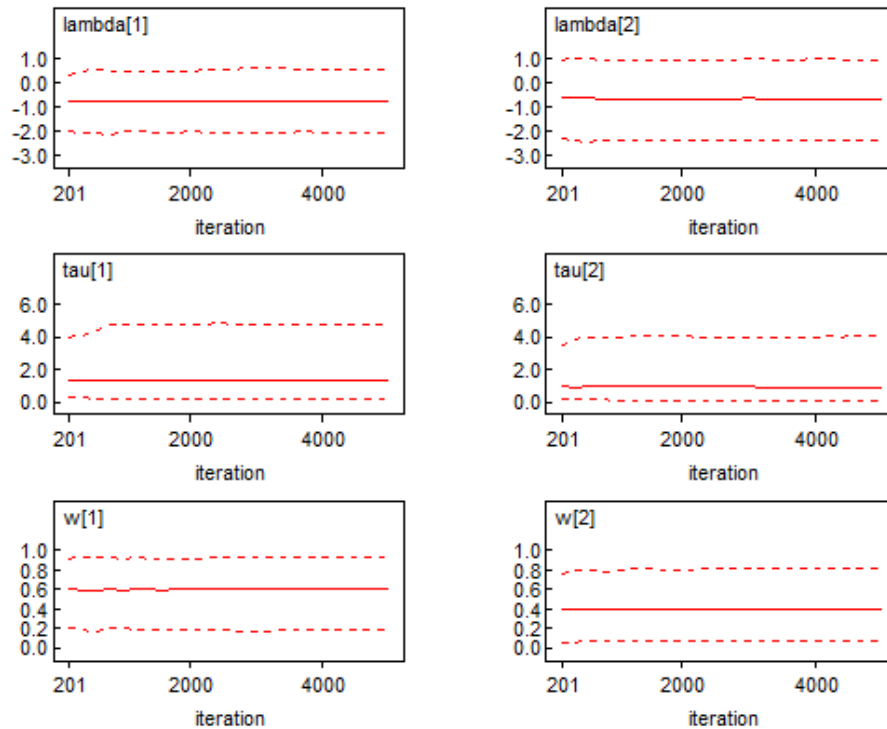


## Lampiran 6. (Lanjutan)



**Gambar L.5.56.** *History* Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534

## Lampiran 6. (Lanjutan)



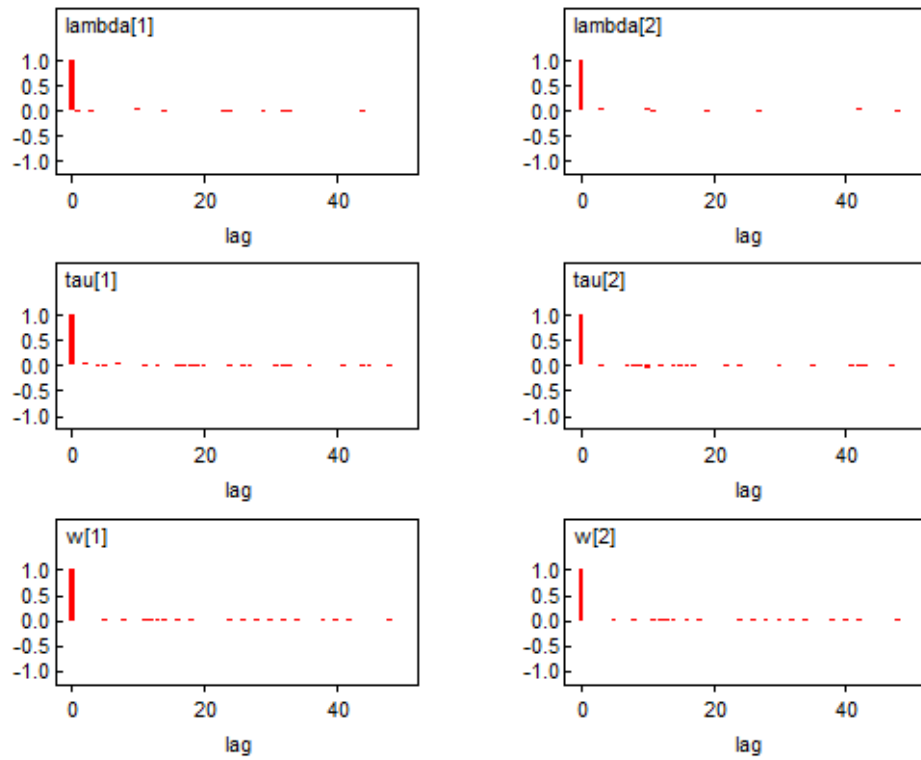
**Gambar L.5.57.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0534

### 15. ID Gen CA0702 (Pembangkitan Parameter 5.000 Kali)

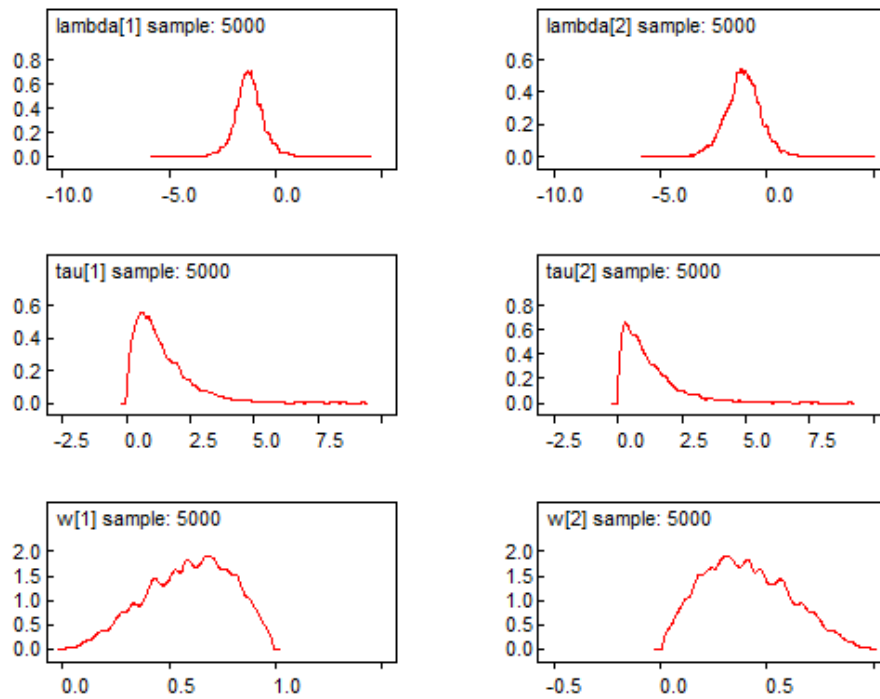
**Tabel L.5.15.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA0702

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-1,221	0,6762	0,0083	-2,551	-1,243	0,1635	1	5.000
lambda[2]	-1,116	0,8366	0,0112	-2,831	-1,108	0,5234	1	5.000
tau[1]	1,362	1,051	0,0151	0,1367	1,091	4,082	1	5.000
tau[2]	1,202	1,039	0,0134	0,0824	0,9082	4,062	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

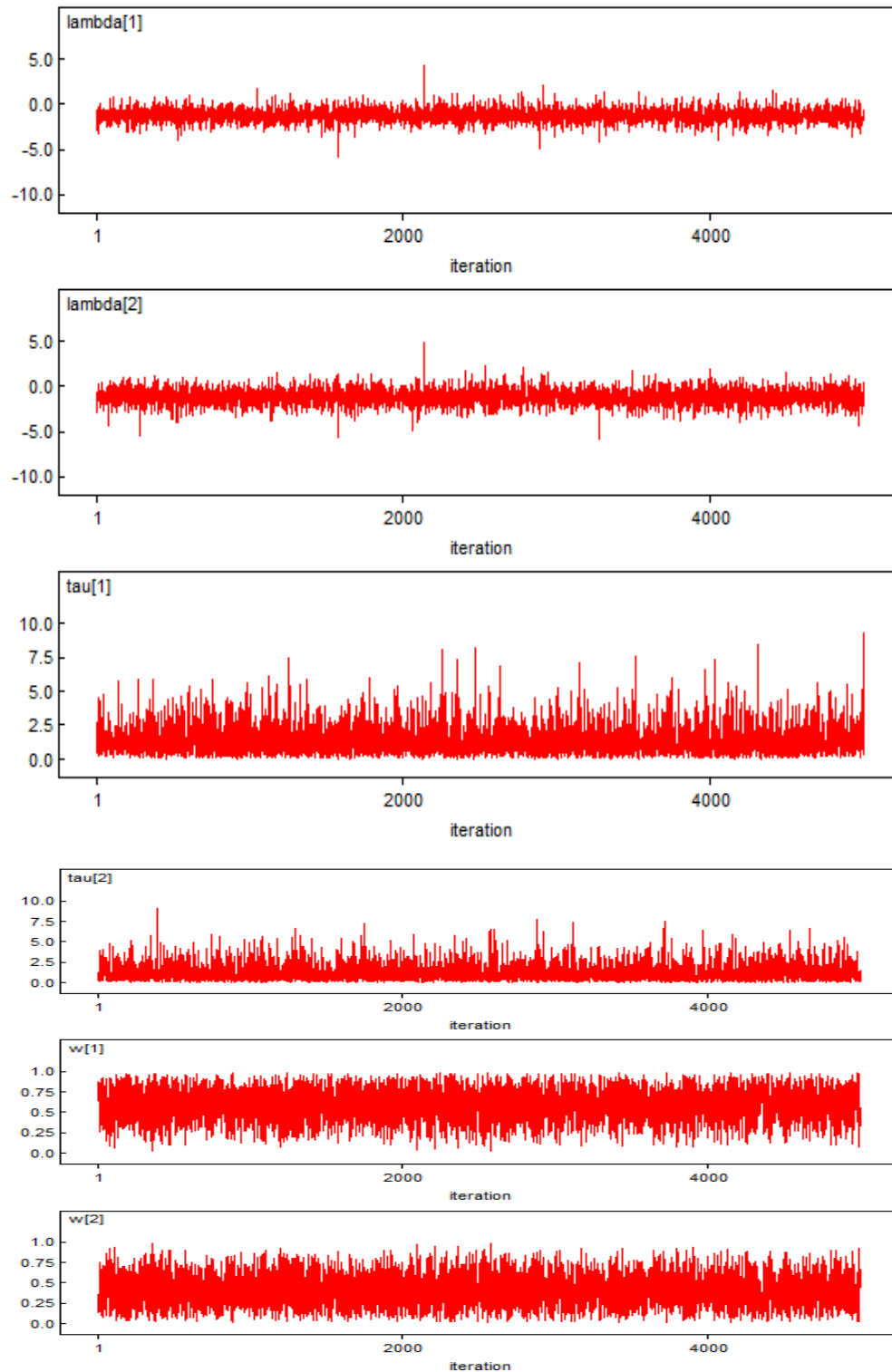


**Gambar L.5.58.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702



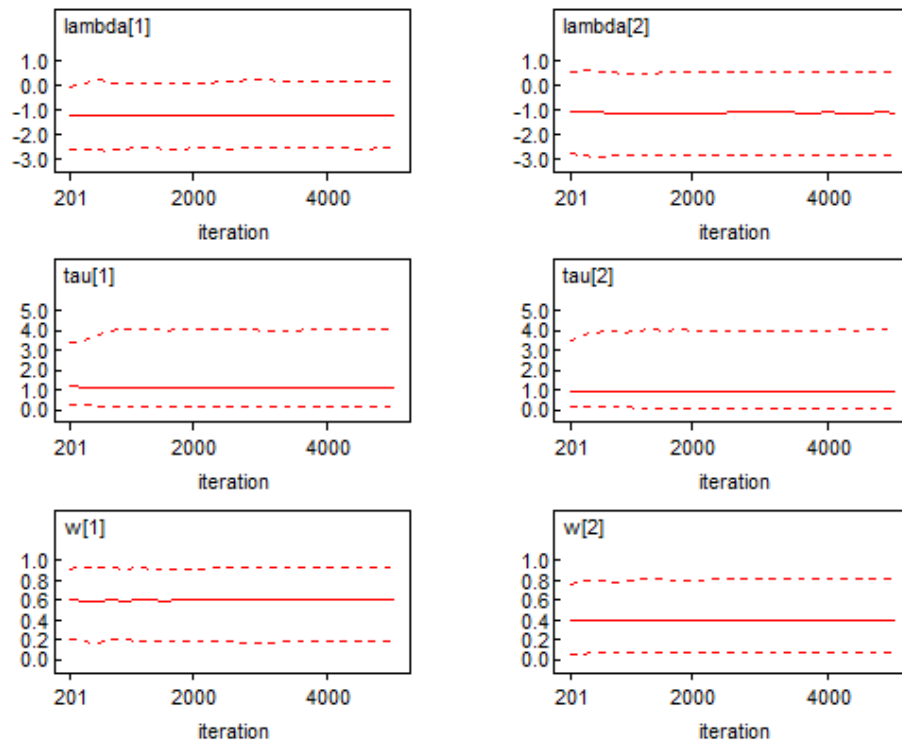
**Gambar L.5.59.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702

## Lampiran 6. (Lanjutan)



**Gambar L.5.60.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702*

## Lampiran 6. (Lanjutan)



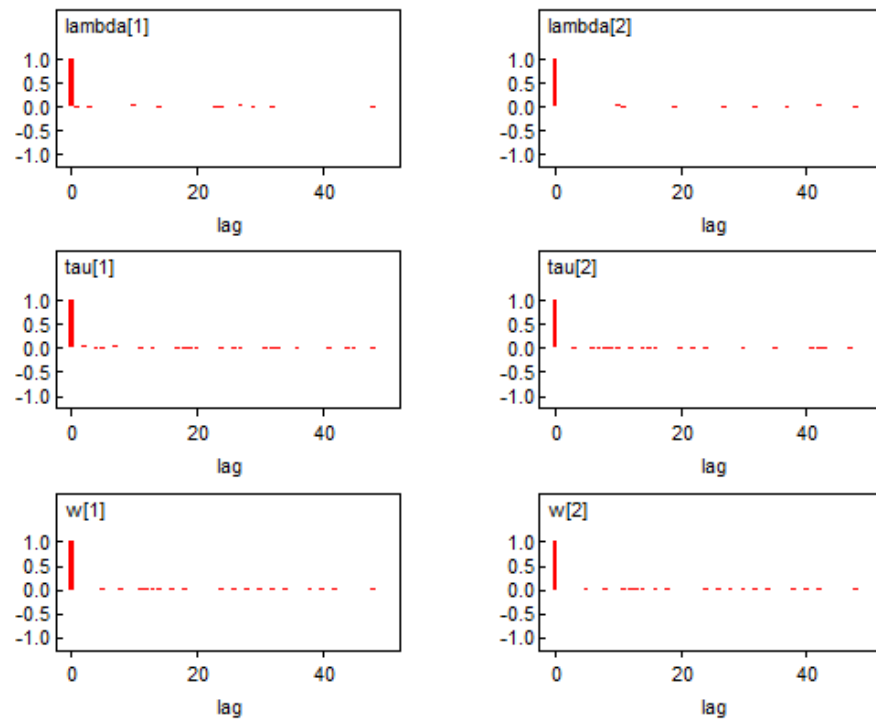
**Gambar L.5.61.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0702

## 16. ID Gen CA1120 (Pembangkitan Parameter 5.000 Kali)

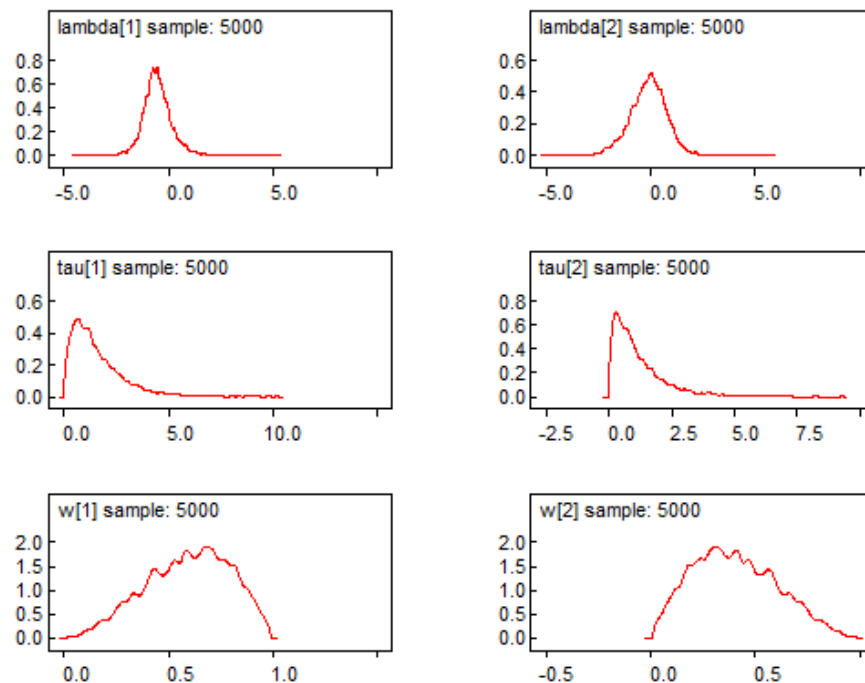
**Tabel L.5.16.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA1120

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,5258	0,6738	0,008	-1,755	-0,5673	0,9563	1	5.000
lambda[2]	-0,1296	0,8781	0,0116	-2,007	-0,0822	1,499	1	5.000
tau[1]	1,556	1,227	0,0181	0,1452	1,234	4,774	1	5.000
tau[2]	1,133	1,015	0,0133	0,0755	0,8405	3,916	1	5.000
w[1]	0,5959	0,1997	0,0029	0,1837	0,6107	0,929	1	5.000
w[2]	0,4041	0,1997	0,0029	0,0713	0,3893	0,8169	1	5.000

## Lampiran 6. (Lanjutan)

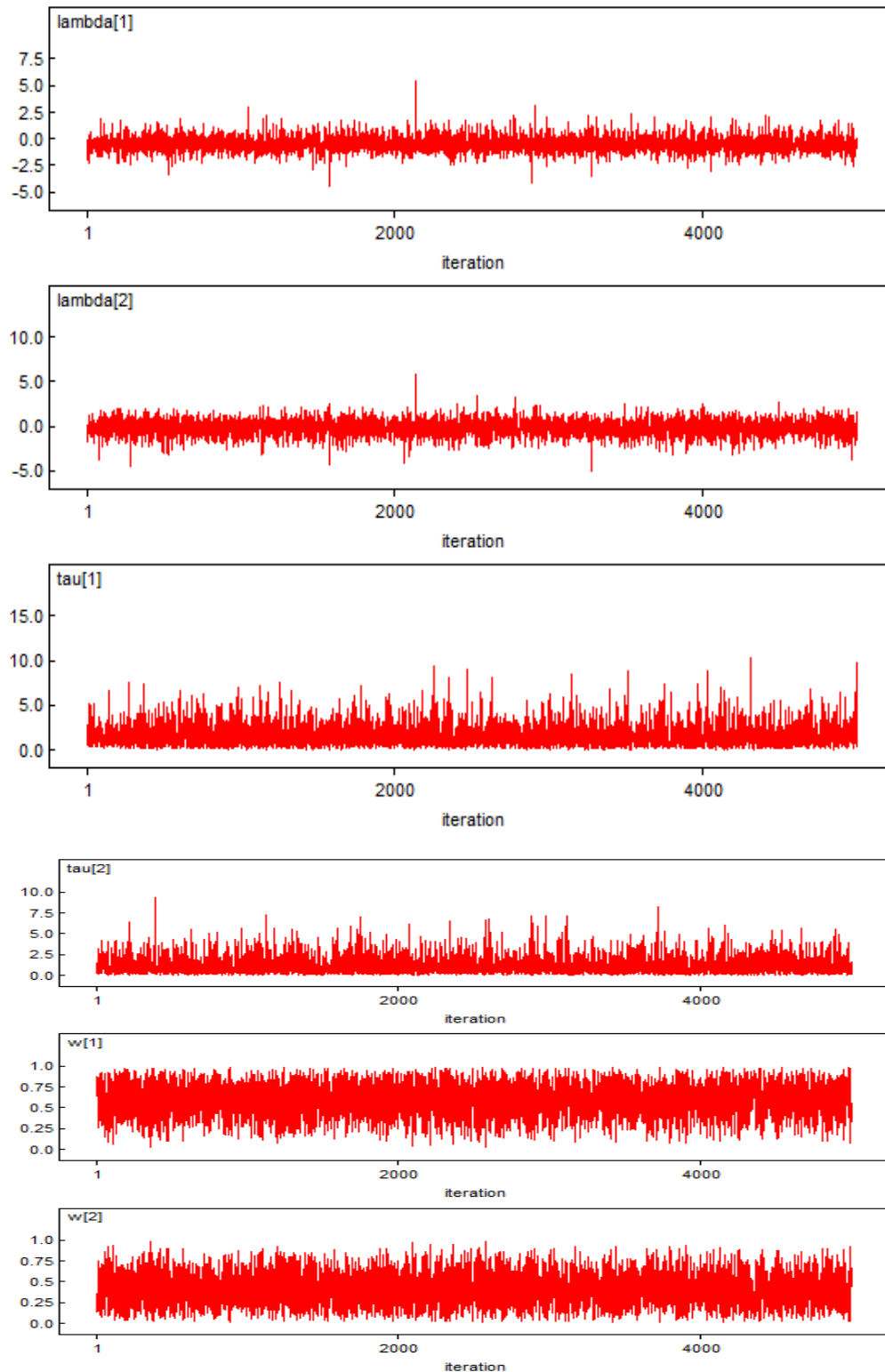


**Gambar L.5.62.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120



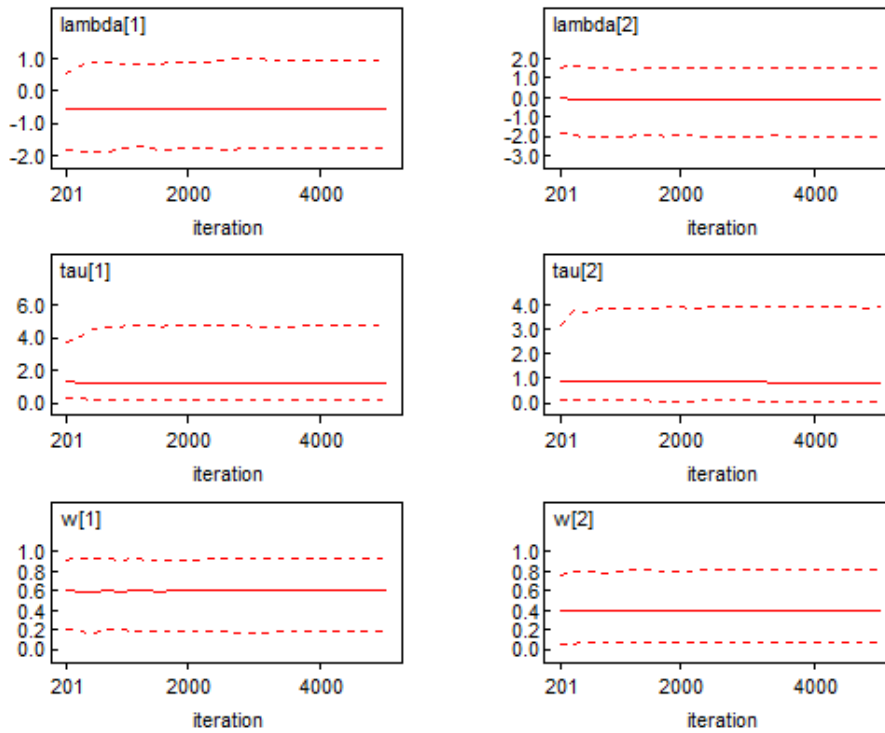
**Gambar L.5.63.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120

## Lampiran 6. (Lanjutan)



**Gambar L.5.64.** *History Data Bangkitan* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120

## Lampiran 6. (Lanjutan)



**Gambar L.5.65.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA1120

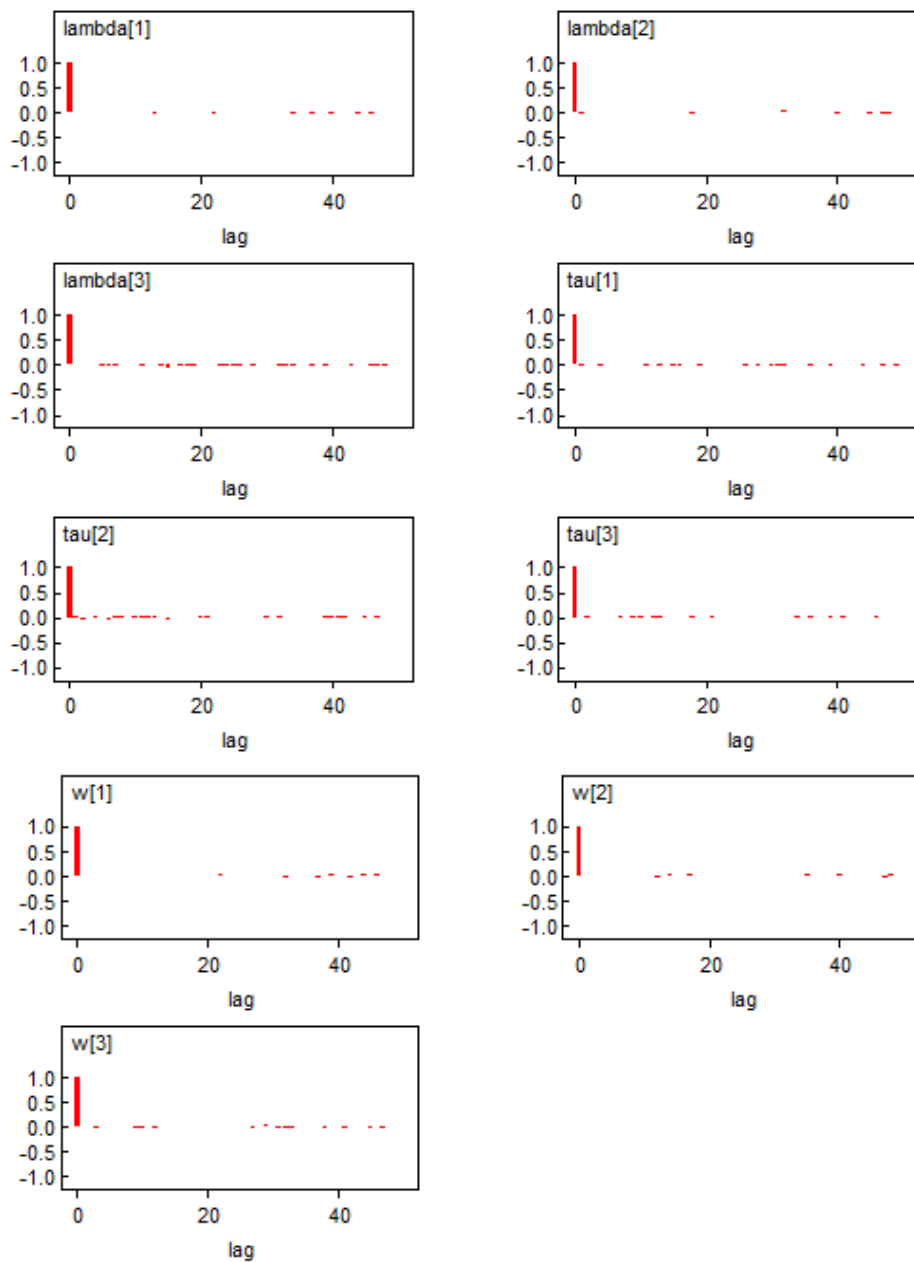
### 17. ID Gen U001 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.5.17.** Estimasi Parameter Model BMM-Normal Tiga Komponen untuk ID Gen U001

Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,7846	0,232	0,0033	-1,264	-0,7744	-0,3417	1	5.000
lambda[2]	-0,4532	0,2061	0,0029	-0,851	-0,4539	-0,0521	1	5.000
lambda[3]	0,0988	0,2445	0,003	-0,348	0,0910	0,6024	1	5.000
tau[1]	12	0,8091	0,0105	10,45	11,98	13,64	1	5.000
tau[2]	15	0,8277	0,0113	13,41	14,99	16,68	1	5.000
tau[3]	15,01	0,8141	0,0115	13,44	15	16,63	1	5.000
w[1]	0,335	0,1768	0,0021	0,053	0,3173	0,7098	1	5.000
w[2]	0,3303	0,1771	0,0025	0,0535	0,3121	0,7198	1	5.000
w[3]	0,3347	0,1781	0,0022	0,0552	0,3166	0,7204	1	5.000

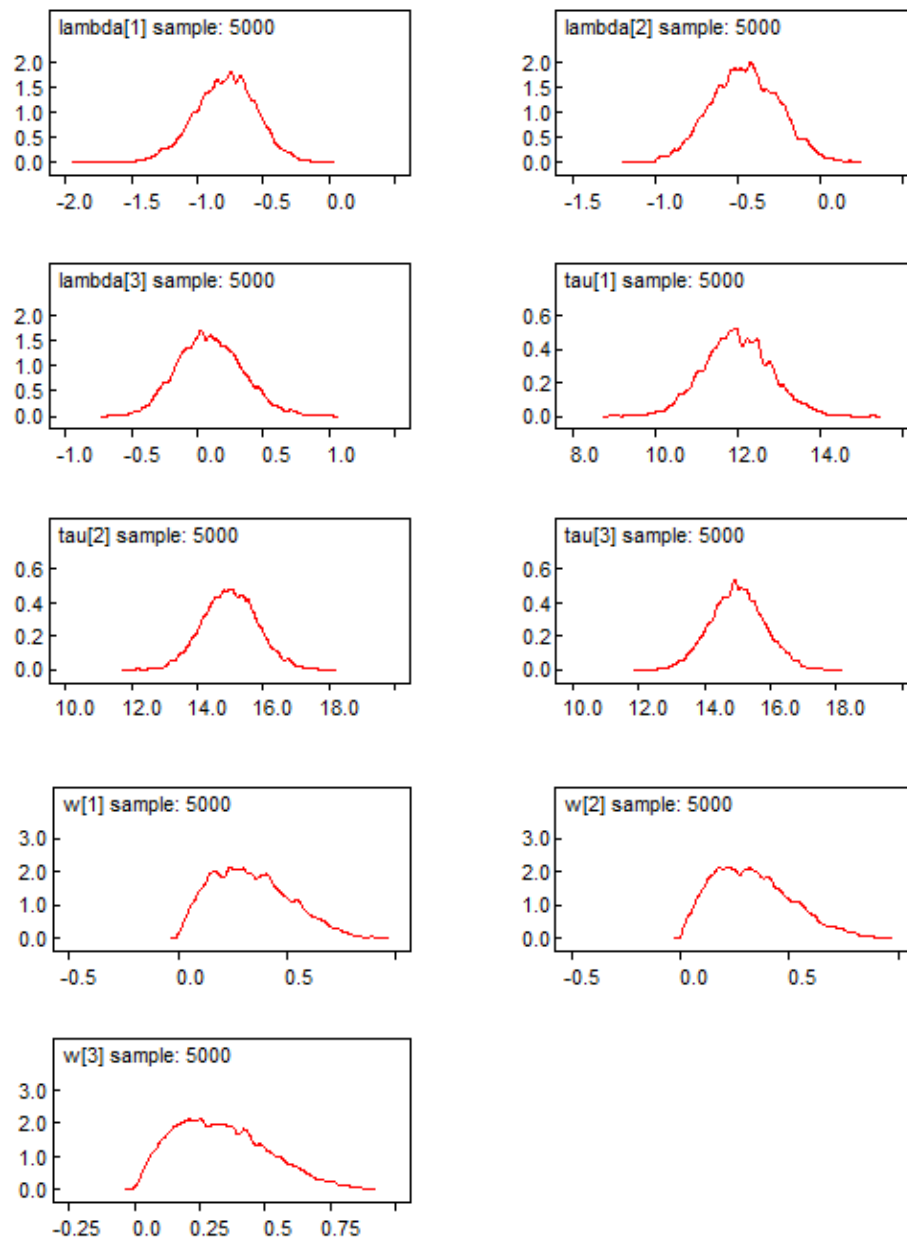


## Lampiran 6. (Lanjutan)



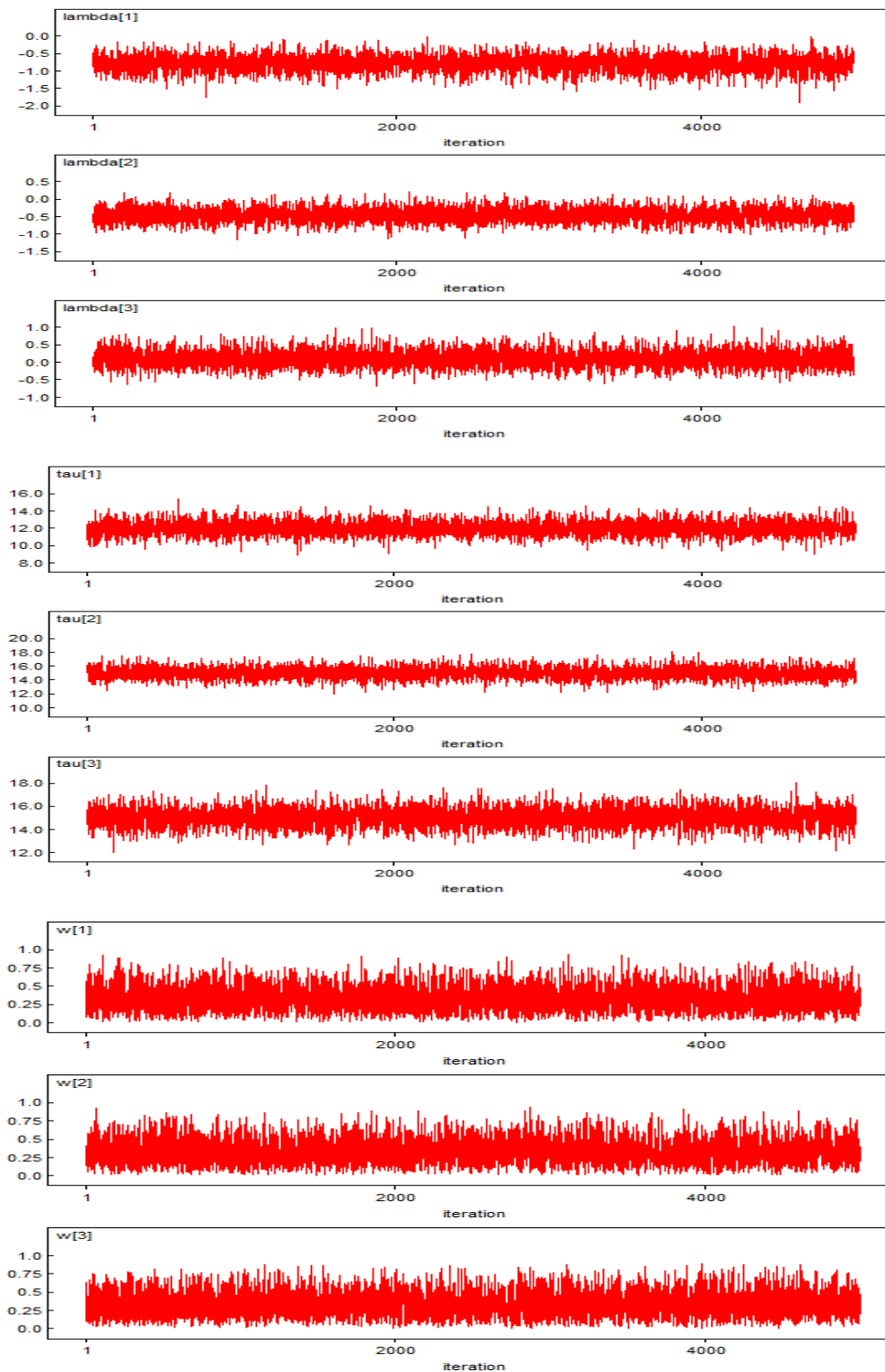
**Gambar L.5.66.** Fungsi Autokorelasi setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

## Lampiran 6. (Lanjutan)



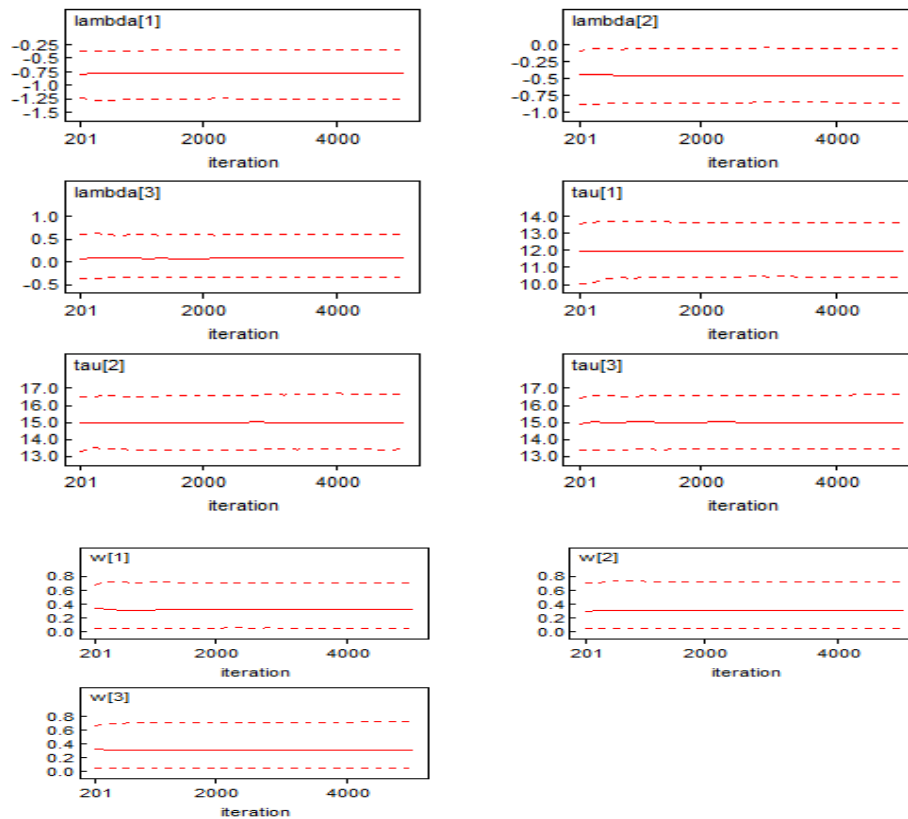
**Gambar L.5.67.** Densitas Kernel setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

## Lampiran 6. (Lanjutan)



**Gambar L.5.68.** *History* Data Bangkitan setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

## Lampiran 6. (Lanjutan)



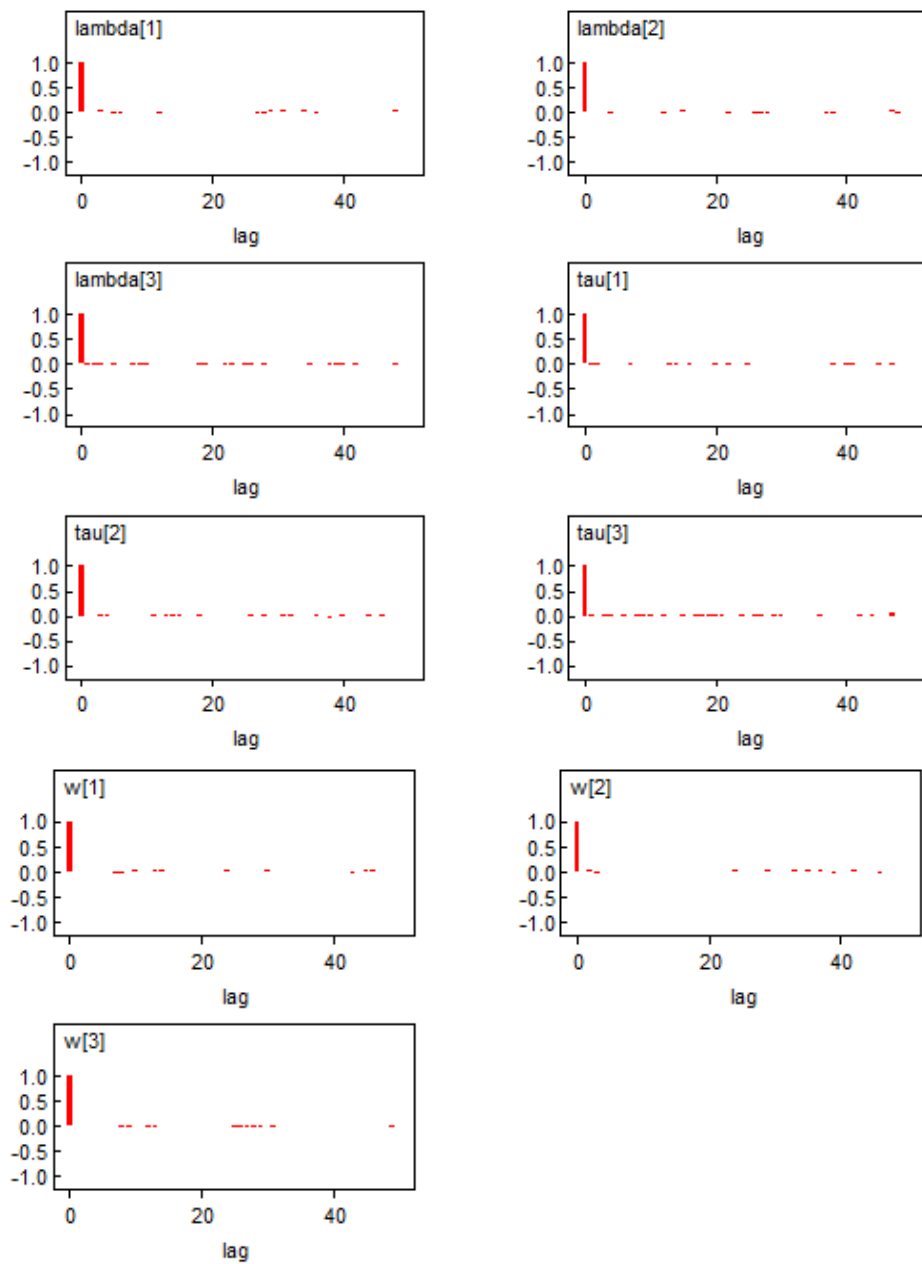
**Gambar L.5.69.** *Quantiles* setiap Parameter BMM-Normal Tiga Komponen untuk ID Gen U001

18. ID Gen CA0810 (Pembangkitan Parameter 5.000 Kali)

**Tabel L.5.18.** Estimasi Parameter Model BMM-Normal Dua Komponen untuk ID Gen CA0810

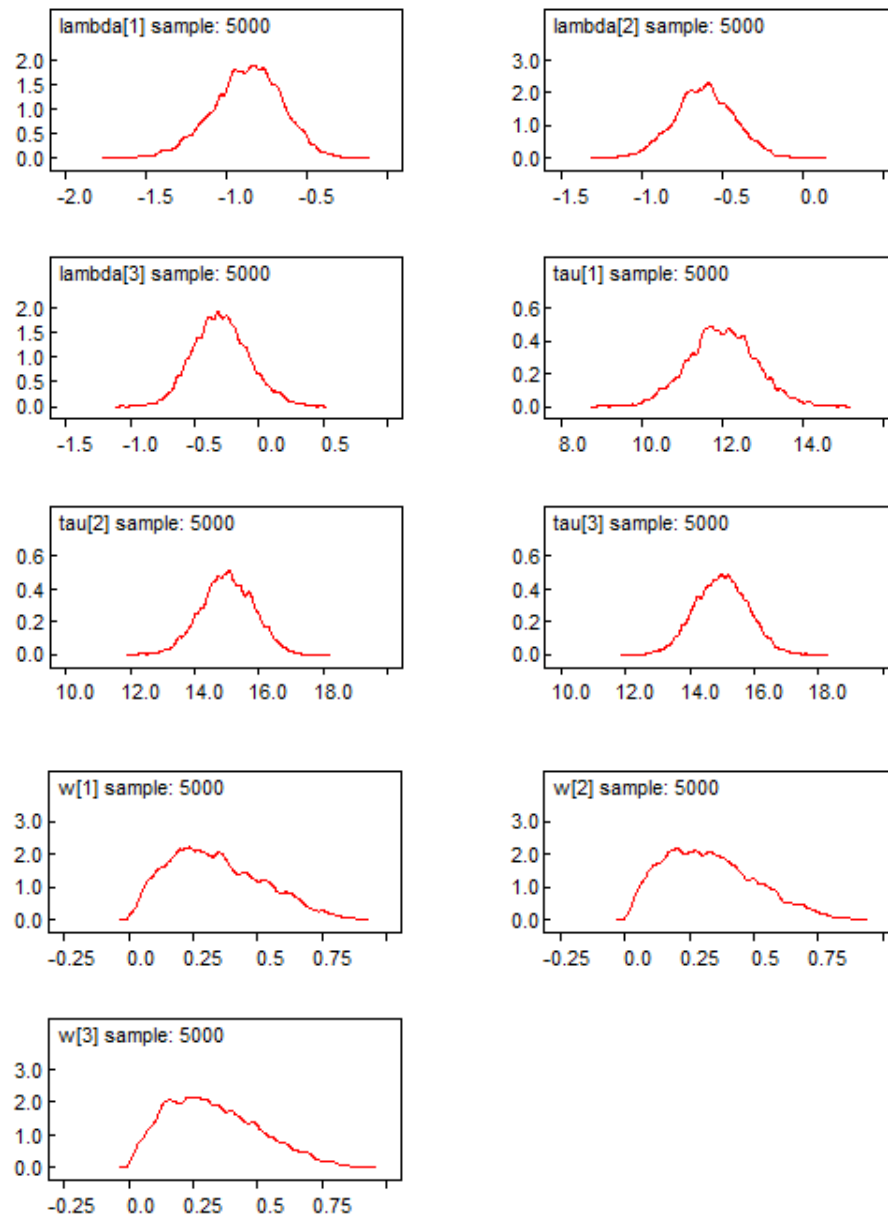
Parameter	Rata-Rata	Simpangan Baku	MC Error	2,50%	Median	97,50%	Start	Sample
lambda[1]	-0,871	0,2152	0,0029	-1,314	-0,8607	-0,4716	1	5.000
lambda[2]	-0,609	0,1875	0,0024	-0,975	-0,6087	-0,2448	1	5.000
lambda[3]	-0,293	0,2161	0,0025	-0,694	-0,3012	0,1607	1	5.000
tau[1]	12,01	0,8268	0,0105	10,41	12	13,65	1	5.000
tau[2]	15,02	0,8104	0,0102	13,43	15,02	16,58	1	5.000
tau[3]	15	0,8099	0,0143	13,45	15	16,56	1	5.000
w[1]	0,3389	0,1804	0,0027	0,0568	0,316	0,7236	1	5.000
w[2]	0,33	0,1761	0,0026	0,0552	0,3127	0,7133	1	5.000
w[3]	0,3311	0,1792	0,0025	0,0502	0,3084	0,7181	1	5.000

## Lampiran 6. (Lanjutan)



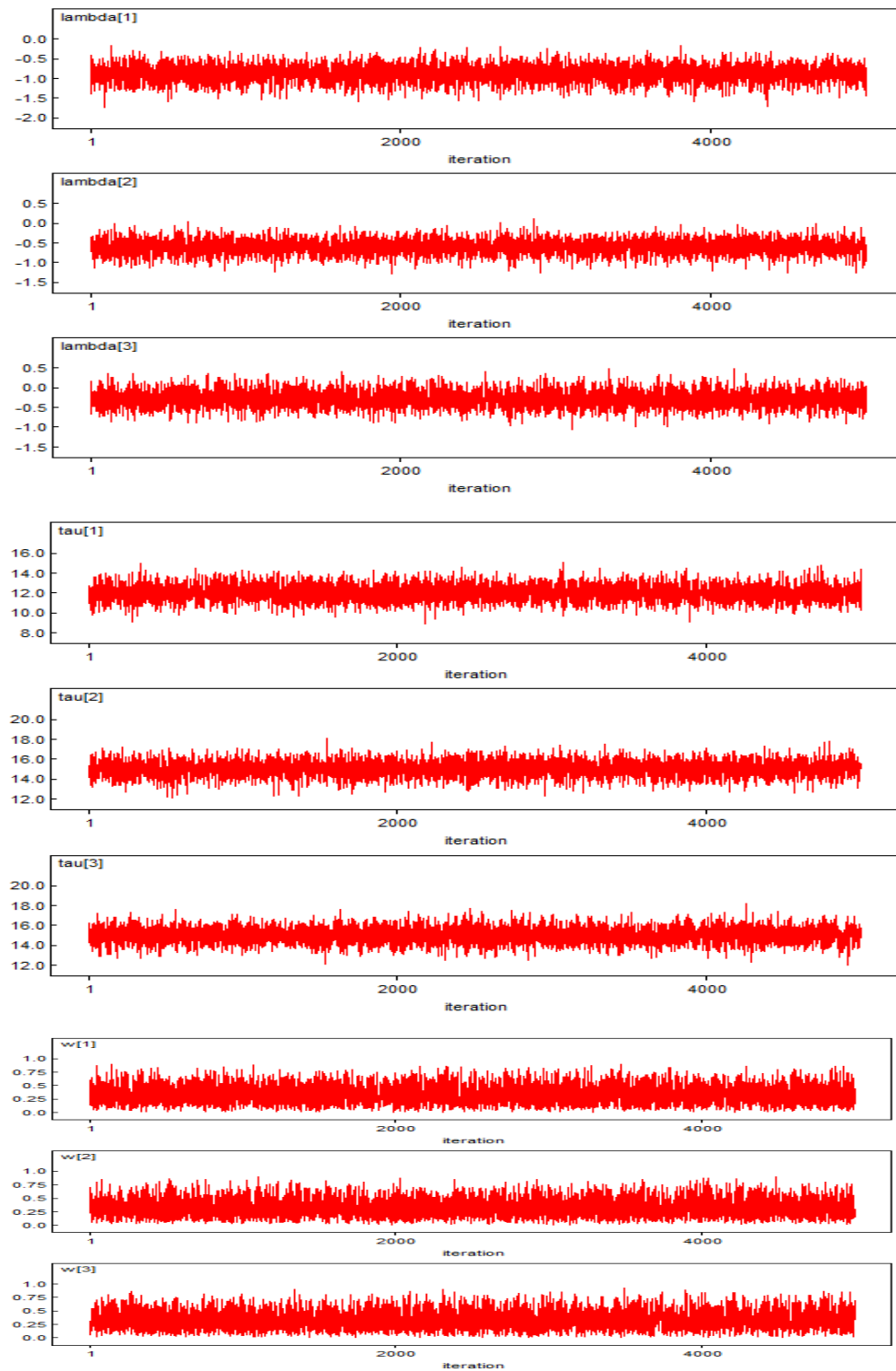
**Gambar L.5.70.** Fungsi Autokorelasi setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0810

## Lampiran 6. (Lanjutan)



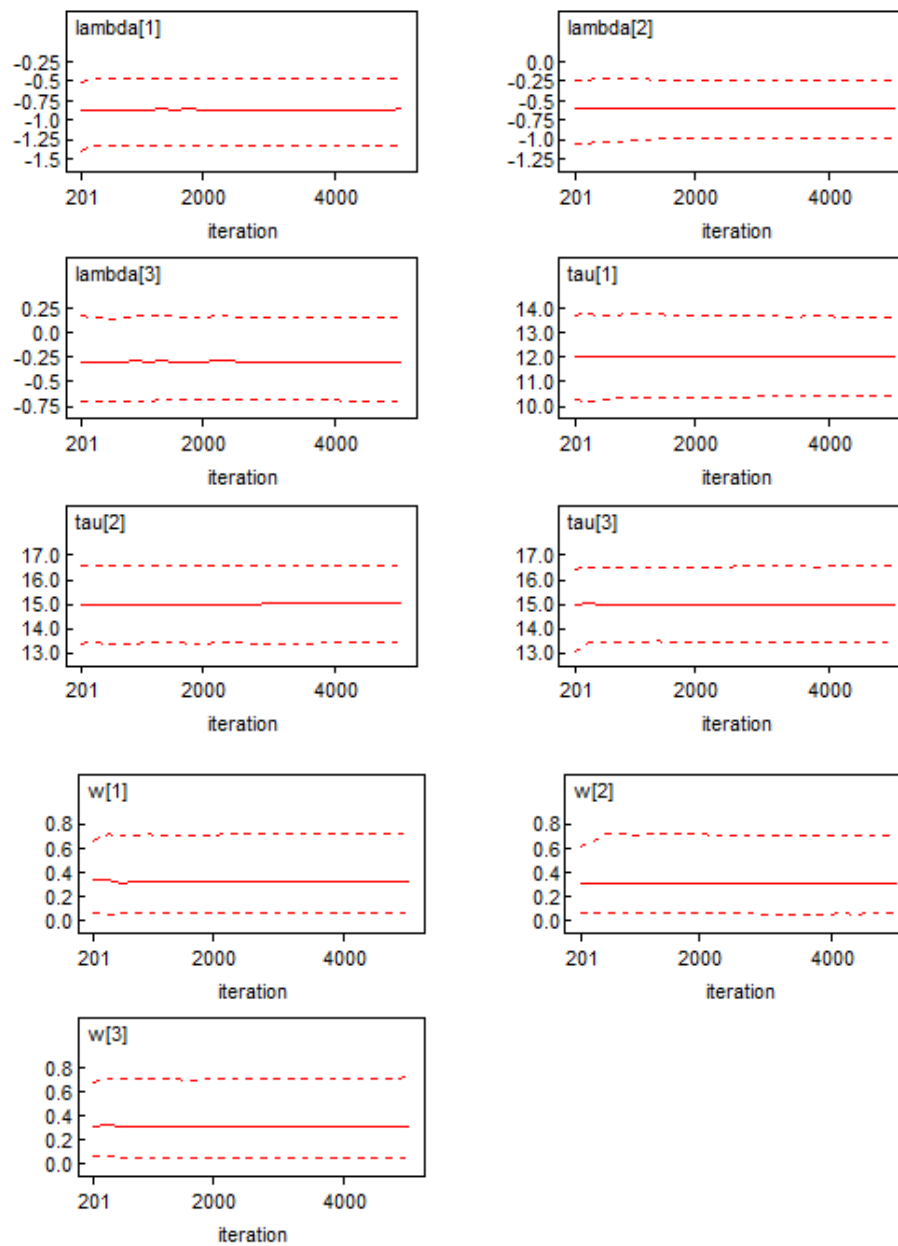
**Gambar L.5.71.** Densitas Kernel setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0810

## Lampiran 6. (Lanjutan)



**Gambar L.5.72.** *History Data Bangkitan setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0810*

## Lampiran 6. (Lanjutan)



**Gambar L.5.73.** *Quantiles* setiap Parameter BMM-Normal Dua Komponen untuk ID Gen CA0810



*Halaman ini sengaja dikosongkan*

## Lampiran 7. Code Program R dari Feng (2013)

### Program Utama:

```
function (y, nsweep, kmax, k, w, mu, sigma2, Z, delta = 1, xi = NULL,
  kappa = NULL, alpha = 2, beta = NULL, g = 0.2, h = NULL,
  verbose = TRUE)
{
  if (nsweep <= 0)
    stop("The number of sweeps has to be positive.")
  if (kmax < 0 || k < 0)
    stop("The number of components have to be positive.")
  if (kmax < k)
    stop("The maximum number of components allowed is larger than\n      the initial value of components.")
  if (length(w) != k) {
    w <- rep(w, len = k)
    warning("The length of 'w' was not equal to k and\n      was forced to be k by being cut off or recycled.")
  }
  if (any(w <= 0))
    stop()
  if (sum(w) != 1) {
  }
  if (length(mu) != k) {
    mu <- rep(mu, len = k)
    warning("The length of 'mu' was not equal to k and\n      was forced to be k by being cut off or recycled.")
  }
  if (length(sigma2) != k) {
    sigma2 <- rep(sigma2, len = k)
    warning("The length of 'sigma2' was not equal to k and\n      was forced to be k by being cut off or recycled.")
  }
  if (any(sigma2 <= 0))
    stop()
  if (length(Z) != length(y)) {
    Z <- rep(Z, len = length(y))
    warning("The length of 'Z' was not equal to the length of 'y' and\n      was forced to be equal by being cut off or recycled.")
  }
  R <- diff(range(y))
  if (is.null(xi))
    xi <- median(y)
  if (is.null(kappa))
    kappa <- 1/R^2
  if (is.null(alpha))
    alpha <- 2
  if (is.null(g))
    g <- 0.2
  if (is.null(h))
    h <- 10/R^2
  if (is.null(beta))
    beta <- rgamma(1, shape = g, rate = h)
  n = length(y)
  b <- rep(0.5, kmax)
  b[kmax] <- 0
  d <- rep(0.5, kmax)
  d[1] <- 0
  k.save <- rep(0, nsweep)
  w.save <- mu.save <- sigma2.save <- vector("list", nsweep)
  Z.save <- matrix(0, nrow = n, ncol = nsweep)
  for (i in 1:nsweep) {
    k <- length(mu)
    Z.expand <- do.call(cbind, lapply(1:k, function(i) ifelse(Z ==
      i, 1, 0)))
    Nj <- colSums(Z.expand)
    w <- rdirichlet(1, delta + Nj)
    sumYj <- colSums(y * Z.expand)
    precision <- Nj/sigma2 + kappa
    mean <- (sumYj/sigma2 + kappa * xi)/precision
    mu.new <- rnorm(k, mean = mean, sd = sqrt(1/precision))
    if (all(order(mu.new) == 1:k)) {
      mu <- mu.new
    }
  }
}
```

## Lampiran 7. (Lanjutan)

```
Diff2j <- outer(y, mu, `^`)^2
sumDiff2j <- colSums(Diff2j * Z.expand)
sigma2 <- rgamma(k, shape = alpha + Nj/2, rate = beta +
  sumDiff2j/2)
sigma2 <- 1/sigma2
p <- do.call(cbind, lapply(1:k, function(i) w[i] * dnorm(y,
  mu[i], sqrt(sigma2[i])))))
s <- rowSums(p)
p[which(s == 0 | !is.finite(s)), ] <- 1
p <- p/rowSums(p)
if (ncol(p) > 1) {
  Z <- rMultinom(p)
}
else {
  Z <- rep(1, n)
}
beta <- rgamma(1, shape = g + k * alpha, rate = h + sum(1/sigma2))
action <- decideScBd(k, kmax)
if (action == 1) {
  split.results <- split(y, k, w, mu, sigma2, Z, b,
    d, delta, xi, kappa, alpha, beta)
  w <- split.results$w
  mu <- split.results$mu
  sigma2 <- split.results$sigma2
  Z <- split.results$Z
  k <- length(w)
}
else {
  combine.results <- combine(y, k, w, mu, sigma2, Z,
    b, d, delta, xi, kappa, alpha, beta)
  w <- combine.results$w
  mu <- combine.results$mu
  sigma2 <- combine.results$sigma2
  Z <- combine.results$Z
  k <- length(w)
}
action <- decideScBd(k, kmax)
if (action == 1) {
  birth.results <- birth(n, k, w, mu, sigma2, Z, delta,
    xi, kappa, alpha, beta, b, d)
  w <- birth.results$w
  mu <- birth.results$mu
  sigma2 <- birth.results$sigma2
  Z <- birth.results$Z
  k <- length(w)
}
else {
  death.results <- death(n, k, w, mu, sigma2, Z, delta,
    xi, kappa, alpha, beta, b, d)
  w <- death.results$w
  mu <- death.results$mu
  sigma2 <- death.results$sigma2
  Z <- death.results$Z
  k <- length(w)
}
k.save[i] <- k
w.save[[i]] <- w
mu.save[[i]] <- mu
sigma2.save[[i]] <- sigma2
Z.save[, i] <- Z
if (verbose && i%1000 == 0) {
  cat(paste(i, " sweeps", " have finished.\n", sep = ""))
}
}
list(k.save = k.save, w.save = w.save, mu.save = mu.save,
  sigma2.save = sigma2.save, Z.save = Z.save)
}
<environment: namespace:miscF>
```

## Lampiran 7. (Lanjutan)

### Program untuk menetapkan *split-merge* dan *birth-death*.

```
#decide whether to split-combine and birth-death
decideScBd <- function(k, kmax){
  if(k==1)
    action <- 1
  else{
    if(k==kmax)
      action <- 0
    else
      action <- ifelse(runif(1) < 0.5, 1, 0)
  }
}

#get the probabilities for z in the split and combine cases
getP <- function(yjstar, muj1, muj2, sigma2j1, sigma2j2, wj1, wj2){
  #p <- exp(cbind(-(yjstar-muj1)^2 / (2*sigma2j1),
  #              -(yjstar-muj2)^2 / (2*sigma2j2)))
  #p[,1] <- p[,1] * wj1 / sqrt(sigma2j1)
  #p[,2] <- p[,2] * wj2 / sqrt(sigma2j2)
  p <- cbind(dnorm(yjstar, muj1, sqrt(sigma2j1)),
             dnorm(yjstar, muj2, sqrt(sigma2j2)))
  p[,1] <- p[,1] * wj1
  p[,2] <- p[,2] * wj2
  s <- rowSums(p)
  p[which(s == 0 | !is.finite(s)),] <- 1
  p / rowSums(p)
}

#get acceptance probabilities for split-combine
getAsc <- function(yjstar, j1.pos, j2.pos, k,
                  wj1, wj2, wjstar,
                  muj1, muj2, mujstar,
                  sigma2j1, sigma2j2, sigma2jstar,
                  delta, xi, kappa, alpha, beta,
                  p, b, d, u1, u2, u3)
){
#browser()
#calculate acceptance probability
## take log first then transform to exponential. Cause there are
## multiplication of larger number of densities and so on.

#likelihood ratio
log.lr <- sum(dnorm(yjstar[j1.pos], mean=muj1, sd=sqrt(sigma2j1), log=TRUE)) +
  sum(dnorm(yjstar[j2.pos], mean=muj2, sd=sqrt(sigma2j2), log=TRUE)) -
  sum(dnorm(yjstar, mean=mujstar, sd=sqrt(sigma2jstar), log=TRUE))
l1 <- length(j1.pos)
l2 <- length(j2.pos)

#term1-3: ratio between two states
#term1 includes various ratios
log.term1 <- log.lr + #likelihood ratio
  log(1) + #ratio of p(k+1)/p(k)
  (delta-l1+l2)*log(wj1) + (delta-l1+l2)*log(wj2) - (delta-l1+l2)*log(wjstar) - #ratio of z
  lbeta(delta, k*delta) #ratio of w
  # (lgamma(k*delta+n) + lgamma(delta+l1) + lgamma(delta+l2) - lgamma((k+1)*delta + n) - lgamma(delta+l1+l2))

#term2 is the ratio of mu
log.term2 <- log(k+1) +
  0.5*log(kappa/(2*pi)) -
  0.5*kappa*((muj1-xi)^2 + (muj2-xi)^2 - (mujstar-xi)^2)

#term3 is the ratio of sigma
log.term3 <- alpha*log(beta) - lgamma(alpha) +
  (-alpha-1)*(log(sigma2j1) + log(sigma2j2) - log(sigma2jstar)) -
  beta*(1/sigma2j1 + 1/sigma2j2 - 1/sigma2jstar)

#term4 is the transform ratio between two states
log.Palloc <- sum(log(p[,1][j1.pos])) + sum(log(p[,2][j2.pos]))
log.term4 <- log(d[k+1]) - log(b[k]) - log.Palloc -
```

## Lampiran 7. (Lanjutan)

```

dbeta(u1, 2, 2, log=TRUE) - dbeta(u2, 2, 2, log=TRUE) - dbeta(u3, 1, 1, log=TRUE)

#term5 is the Jacobian
log.term5 <- log(wjstar) + log(abs(muj1-muj2)) + log(sigma2j1) + log(sigma2j2) -
  log(u2) - log(1-u2^2) - log(u3) - log(1-u3) - log(sigma2jstar)

exp(log.term1 + log.term2 + log.term3 + log.term4 + log.term5)
}

split <- function(y, k, w, mu, sigma2, Z, b, d, delta, xi, kappa, alpha, beta){

  #choose a component to split
  jstar <- sample(1:k, 1)
  #generate intermediate parameters
  u1 <- rbeta(1, 2, 2)
  u2 <- rbeta(1, 2, 2)
  u3 <- rbeta(1, 1, 1)

  #generate two new ws
  wjstar <- w[jstar]
  wj1 <- wjstar * u1
  wj2 <- wjstar * (1-u1)

  #generate two new mus
  mujstar <- mu[jstar]
  sigma2jstar <- sigma2[jstar]
  muj1 <- mujstar - u2*sqrt(sigma2jstar)*sqrt(wj2/wj1)
  muj2 <- mujstar + u2*sqrt(sigma2jstar)*sqrt(wj1/wj2)

  #check order of mu
  newmu.part <- c(mu[jstar-1], muj1, muj2, mu[jstar+1])
  if(all(order(newmu.part) == 1:length(newmu.part))){
    #generate two new sigma2s
    sigma2j1 <- u3 * (1-u2^2) * sigma2jstar * wjstar / wj1
    sigma2j2 <- (1-u3) * (1-u2^2) * sigma2jstar * wjstar / wj2

    #allocate z_i=jstar
    jstar.pos <- which(Z==jstar)
    if(length(jstar.pos) > 0){
      yjstar <- y[jstar.pos]
      p <- getP(yjstar, muj1, muj2, sigma2j1, sigma2j2, wj1, wj2)
      zj12 <- rMultinom(p)
      j1.pos <- which(zj12==1)
      j2.pos <- which(zj12==2)

      A <- getAsc(yjstar, j1.pos, j2.pos, k,
        wj1, wj2, wjstar,
        muj1, muj2, mujstar,
        sigma2j1, sigma2j2, sigma2jstar,
        delta, xi, kappa, alpha, beta,
        p, b, d, u1, u2, u3)

      if(runif(1) < min(1, A)){
        indicator <- rep(0,k)
        indicator[jstar] <- 1
        indicator[which((1:k) > jstar)] <- 2
        ind0 <- which(indicator==0)
        ind2 <- which(indicator==2)

        #generate new w vector
        w <- c(w[ind0], wj1, wj2, w[ind2])
        #generate new mu vector
        mu <- c(mu[ind0], muj1, muj2, mu[ind2])
        #generate new sigma2 vector
        sigma2 <- c(sigma2[ind0], sigma2j1, sigma2j2, sigma2[ind2])
        #generate new Z matrix
        larger <- which(Z > jstar)
        Z[larger] <- Z[larger] + 1
        Z[jstar.pos[j1.pos]] <- jstar
        Z[jstar.pos[j2.pos]] <- jstar + 1
      }
    }
  }
}

```

## Lampiran 7. (Lanjutan)

```

    }
  }
}
list(w=w, mu=mu, sigma2=sigma2, Z=Z)
}
combine <- function(y, k, w, mu, sigma2, Z, b, d, delta, xi, kappa, alpha, beta){

  #choose a pair of components to combine
  j1 <- sample(1:(k-1), 1)
  j2 <- j1 + 1

  #generate new parameters
  wj1 <- w[j1]
  wj2 <- w[j2]
  muj1 <- mu[j1]
  muj2 <- mu[j2]
  sigma2j1 <- sigma2[j1]
  sigma2j2 <- sigma2[j2]
  wjstar <- w[j1] + w[j2]
  mujstar <- (wj1*muj1 + wj2*muj2) / wjstar

  #Note the sigma2jstar is derived from the split and is different than
  # that from eqn (10)
  #sigma2jstar <- (wj1*(muj1^2+sigma2j1) + wj2*(muj2^2+sigma2j2)) /
  #      wjstar - mujstar^2
  sigma2jstar = wj1*wj2*((muj1-muj2)/wjstar)**2 +
    (wj1*sigma2j1+wj2*sigma2j2)/wjstar

  #calculate acceptance probability
  #likelihood ratio
  jstar.pos <- which(Z==j1 | Z==j2)
  j1.pos <- match(which(Z==j1), jstar.pos)
  j2.pos <- match(which(Z==j2), jstar.pos)
  yjstar <- y[jstar.pos]
  p <- getP(yjstar, muj1, muj2, sigma2j1, sigma2j2, wj1, wj2)

  #generate intermediate parameters
  #u1 u2 u3 are derived from split move (equations below eqn (10) on page 739
  #u1 <- rbeta(1, 2, 2)
  #u2 <- rbeta(1, 2, 2)
  #u3 <- rbeta(1, 1, 1)
  u1 <- wj1 / wjstar
  u2 = (muj2-muj1) * sqrt(wj1*wj2/sigma2jstar) / wjstar
    u2 = max(u2, 1e-12)
    u2 = min(u2, 1.0-1e-4)
  u3 = wj1*sigma2j1 / (wj1*sigma2j1+wj2*sigma2j2)

  A <- getAsc(yjstar, j1.pos, j2.pos, k-1,
    wj1, wj2, wjstar,
    muj1, muj2, mujstar,
    sigma2j1, sigma2j2, sigma2jstar,
    delta, xi, kappa, alpha, beta,
    p, b, d, u1, u2, u3)

  if(runif(1) < min(1, 1/A)){

    #browser()
    indicator <- rep(0,k)
    indicator[c(j1,j2)] <- 1
    indicator[which((1:k) > j2)] <- 2
    ind0 <- which(indicator==0)
    ind2 <- which(indicator==2)

    #generate new w vector
    w <- c(w[ind0], wjstar, w[ind2])
    #generate new mu vector
    mu <- c(mu[ind0], mujstar, mu[ind2])
    #generate new sigma2 vector
    sigma2 <- c(sigma2[ind0], sigma2jstar, sigma2[ind2])
    #generate new Z matrix

```

## Lampiran 7. (Lanjutan)

```

Z[which(Z==j2)] <- j1
large <- which(Z > j2)
Z[large] <- Z[large] - 1
}
list(w=w, mu=mu, sigma2=sigma2, Z=Z)
}

getAbd <- function(n, k, k0, delta, wjstar, b, d){
  log.term1 <- log(1) - lbeta(k*delta, delta) + (delta-1)*log(wjstar) +
    (n+k*delta-k)*log(1-wjstar) + log(k+1)

  #note that there is an error in the original paper:
  # (1-wjstar)^(k-1) instead of (1-wjstar)^k
  log.term2 <- log(d[k+1]) - log(k0+1) - log(b[k]) - dbeta(wjstar,1,k, log=TRUE) + (k-1)*log(1-wjstar)
  exp(log.term1 + log.term2)
}

birth <- function(n, k, w, mu, sigma2, Z, delta, xi, kappa, alpha, beta, b, d){

  wjstar <- rbeta(1, 1, k)
  k0 <- sum(unlist(lapply(1:k, function(i) sum(Z==i)))==0)
  A <- getAbd(n, k, k0, delta, wjstar, b, d)
  if(runif(1) < min(1, A)){

    mujstar <- rnorm(1, mean=xi, sd=sqrt(1/kappa))
    sigma2jstar <- rgamma(1, shape=alpha, rate=beta)
    sigma2jstar <- 1 / sigma2jstar
    w <- w*(1-wjstar)
    jstar.pos <- which(mu > mujstar)[1]
    if(is.na(jstar.pos)){
      w <- c(w, wjstar)
      mu <- c(mu, mujstar)
      sigma2 <- c(sigma2, sigma2jstar)
    }
    else{
      indicator <- rep(0,k)
      indicator[which((1:k) >= jstar.pos)] <- 1
      ind0 <- which(indicator==0)
      ind1 <- which(indicator==1)
      w <- c(w[ind0], wjstar, w[ind1])
      mu <- c(mu[ind0], mujstar, mu[ind1])
      sigma2 <- c(sigma2[ind0], sigma2jstar, sigma2[ind1])
      larger <- which(Z >= jstar.pos)
      Z[larger] <- Z[larger] + 1
    }
  }
  list(w=w, mu=mu, sigma2=sigma2, Z=Z)
}

death <- function(n, k, w, mu, sigma2, Z, delta, xi, kappa, alpha, beta, b, d){
  d.candidate <- which(unlist(lapply(1:k, function(i) sum(Z==i)==0)))
  if(length(d.candidate) > 0){
    d.pos <- sample(1:length(d.candidate),1)
    d.pos <- d.candidate[d.pos]
    wjstar <- w[d.pos]
    k0 <- length(d.candidate) - 1
    A <- getAbd(n, k-1, k0, delta, wjstar, b, d)
    if(runif(1) < min(1, 1/A)){
      w <- w[-d.pos]
      w <- w / sum(w)
      mu <- mu[-d.pos]
      sigma2 <- sigma2[-d.pos]
      larger <- which(Z > d.pos)
      Z[larger] <- Z[larger] - 1
    }
  }
  list(w=w, mu=mu, sigma2=sigma2, Z=Z)
}

uvnm.rjmcmc <- function(y, nsweep, kmax, k, w, mu, sigma2, Z,
  delta=1, xi=NULL, kappa=NULL, alpha=2,

```

## Lampiran 7. (Lanjutan)

```
beta=NULL, g=0.2, h=NULL, verbose=TRUE){

#Error checking
if(nswEEP <= 0)
  stop("The number of sweeps has to be positive.")
if(kmax < 0 || k < 0)
  stop("The number of components have to be positive.")
if(kmax < k)
  stop("The maximum number of components allowed is larger than
the initial value of components.")
if(length(w) != k){
  w <- rep(w, len=k)
  warning("The length of 'w' was not equal to k and
was forced to be k by being cut off or recycled.")
}
if(any(w <=0))
  stop()
if(sum(w) != 1){

}

if(length(mu) != k){
  mu <- rep(mu, len=k)
  warning("The length of 'mu' was not equal to k and
was forced to be k by being cut off or recycled.")
}
if(length(sigma2) != k){
  sigma2 <- rep(sigma2, len=k)
  warning("The length of 'sigma2' was not equal to k and
was forced to be k by being cut off or recycled.")
}
if(any(sigma2 <= 0))
  stop
if(length(Z) != length(y)){
  Z <- rep(Z, len=length(y))
  warning("The length of 'Z' was not equal to the length of 'y' and
was forced to be equal by being cut off or recycled.")
}

R <- diff(range(y))
if(is.null(xi))
  xi <- median(y)
if(is.null(kappa))
  kappa <- 1/R^2
if(is.null(alpha))
  alpha <- 2
if(is.null(g))
  g <- 0.2
if(is.null(h))
  h <- 10/R^2
if(is.null(beta))
  beta <- rgamma(1, shape=g, rate=h)

n=length(y)

#split probabilities
b <- rep(0.5, kmax)
b[kmax] <- 0
#combine probabilities
d <- rep(0.5, kmax)
d[1] <- 0

k.save <- rep(0, nswEEP)
w.save <- mu.save <- sigma2.save <- vector("list", nswEEP)
Z.save <- matrix(0, nrow=n, ncol=nswEEP)
for(i in 1:nswEEP){
  k <- length(mu)

  #update w|...
  Z.expand <- do.call(cbind, lapply(1:k, function(i) ifelse(Z==i, 1, 0)))
```



## Lampiran 7. (Lanjutan)

```
Nj <- colSums(Z.expand)
w <- rdirichlet(1, delta + Nj)

#update mu|...
sumYj <- colSums(y * Z.expand)
precision <- Nj/sigma2 + kappa
mean <- (sumYj/sigma2 + kappa*xi) / precision
mu.new <- rnorm(k, mean=mean, sd=sqrt(1/precision))
if(all(order(mu.new) == 1:k)){
  mu <- mu.new
}

#update sigma2|...
Diff2j <- outer(y, mu, `^`) ^ 2
sumDiff2j <- colSums(Diff2j * Z.expand)
sigma2 <- rgamma(k, shape=alpha + Nj/2, rate=beta + sumDiff2j/2)
sigma2 <- 1 / sigma2

#update Z
#p <- exp(- Diff2j %*% diag(1/(2*sigma2), nrow=k, ncol=k))
#p <- p %*% diag(w/sqrt(sigma2), nrow=k, ncol=k)
p <- do.call(cbind, lapply(1:k, function(i)
  w[i] * dnorm(y, mu[i], sqrt(sigma2[i]))))
s <- rowSums(p)
p[which(s == 0 | !is.finite(s)),] <- 1
p <- p / rowSums(p)
if(ncol(p) > 1){
  Z <- rMultinom(p)
}
else{
  Z <- rep(1, n)
}

#update beta
beta <- rgamma(1, shape=g + k*alpha, rate=h + sum(1/sigma2))
#combine or split
action <- decideScBd(k, kmax)
if(action==1){
  split.results <- split(y, k, w, mu, sigma2, Z, b, d, delta, xi, kappa, alpha, beta)
  w <- split.results$w
  mu <- split.results$mu
  sigma2 <- split.results$sigma2
  Z <- split.results$Z
  k <- length(w)
}
else{
  combine.results <- combine(y, k, w, mu, sigma2, Z, b, d, delta, xi, kappa, alpha, beta)
  w <- combine.results$w
  mu <- combine.results$mu
  sigma2 <- combine.results$sigma2
  Z <- combine.results$Z
  k <- length(w)
}

#birth-death
action <- decideScBd(k, kmax)
if(action==1){
  birth.results <- birth(n, k, w, mu, sigma2, Z, delta,
    xi, kappa, alpha, beta, b, d)
  w <- birth.results$w
  mu <- birth.results$mu
  sigma2 <- birth.results$sigma2
  Z <- birth.results$Z
  k <- length(w)
}
else{
  death.results <- death(n, k, w, mu, sigma2, Z, delta,
    xi, kappa, alpha, beta, b, d)
  w <- death.results$w
```

## Lampiran 7. (Lanjutan)

```
mu <- death.results$mu
sigma2 <- death.results$sigma2
Z <- death.results$Z
k <- length(w)
}

k.save[i] <- k
w.save[[i]] <- w
mu.save[[i]] <- mu
sigma2.save[[i]] <- sigma2
Z.save[,i] <- Z

if (verbose && i %% 1000 == 0){
  cat(paste(i, " sweeps", " have finished.\n", sep=""))
}

}
list(k.save=k.save, w.save=w.save, mu.save=mu.save,
     sigma2.save=sigma2.save, Z.save=Z.save)
}
```

*Halaman ini sengaja dikosongkan*

## Lampiran 8. Lembar Pernyataan Keaslian Data

### LEMBAR PERNYATAAN KEASLIAN DATA

Saya yang bertanda tangan di bawah ini ;

Nama : Dra. Nunung Harijati, MS, P.hD  
Jabatan : Dosen Universitas Brawijaya Malang  
NIP : 19611105 199002 2 001

Menyatakan bahwa semua data yang saya berikan pada penelitian disertasi yang berjudul **"Bayesian Mixture Model Averaging untuk Mengidentifikasi Perbedaan Ekspresi Gen Percobaan Microarray"** adalah data yang benar dan sesuai dengan kondisi data asli yang saya miliki.

Demikian surat pernyataan ini dibuat dalam keadaan sadar tanpa adanya tekanan dari pihak manapun.

Malang, 17 Mei 2017

Yang menyatakan,



Dra. Nunung Harijati, MS, P.hD  
NIP. 19611105 199002 2 001

*Halaman ini sengaja dikosongkan*

## Lampiran 9. Model Persamaan BMMA dan Persamaan OLS

### 1. Model Persamaan BMMA untuk berbagai ID gen:

$$\begin{aligned}
 f_{\text{BMMA}}(\text{LS0024}) &= \left[ \hat{w}_{\text{LS0024}(1)(1)} N_{\text{LS0024}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0024}(1)(2)} N_{\text{LS0024}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0024}(939)(1)} N_{\text{LS0024}(939)(1)} \left( \hat{\mu}_{(939)(1)}, \hat{\sigma}_{(939)(1)}^2 \right) + \hat{w}_{\text{LS0024}(939)(2)} N_{\text{LS0024}(939)(2)} \left( \hat{\mu}_{(939)(2)}, \hat{\sigma}_{(939)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{LS0616}) &= \left[ \hat{w}_{\text{LS0616}(1)(1)} N_{\text{LS0616}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0616}(1)(2)} N_{\text{LS0616}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0616}(938)(1)} N_{\text{LS0616}(938)(1)} \left( \hat{\mu}_{(938)(1)}, \hat{\sigma}_{(938)(1)}^2 \right) + \hat{w}_{\text{LS0616}(938)(2)} N_{\text{LS0616}(938)(2)} \left( \hat{\mu}_{(938)(2)}, \hat{\sigma}_{(938)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{LS0159}) &= \left[ \hat{w}_{\text{LS0159}(1)(1)} N_{\text{LS0159}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0159}(1)(2)} N_{\text{LS0159}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0159}(941)(1)} N_{\text{LS0159}(941)(1)} \left( \hat{\mu}_{(941)(1)}, \hat{\sigma}_{(941)(1)}^2 \right) + \hat{w}_{\text{LS0159}(941)(2)} N_{\text{LS0159}(941)(2)} \left( \hat{\mu}_{(941)(2)}, \hat{\sigma}_{(941)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{LS0162}) &= \left[ \hat{w}_{\text{LS0162}(1)(1)} N_{\text{LS0162}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0162}(1)(2)} N_{\text{LS0162}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0162}(929)(1)} N_{\text{LS0162}(929)(1)} \left( \hat{\mu}_{(929)(1)}, \hat{\sigma}_{(929)(1)}^2 \right) + \hat{w}_{\text{LS0162}(929)(2)} N_{\text{LS0162}(929)(2)} \left( \hat{\mu}_{(929)(2)}, \hat{\sigma}_{(929)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{LS0185}) &= \left[ \hat{w}_{\text{LS0185}(1)(1)} N_{\text{LS0185}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0185}(1)(2)} N_{\text{LS0185}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0185}(923)(1)} N_{\text{LS0185}(923)(1)} \left( \hat{\mu}_{(923)(1)}, \hat{\sigma}_{(923)(1)}^2 \right) + \hat{w}_{\text{LS0185}(923)(2)} N_{\text{LS0185}(923)(2)} \left( \hat{\mu}_{(923)(2)}, \hat{\sigma}_{(923)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{LS0081}) &= \left[ \hat{w}_{\text{LS0081}(1)(1)} N_{\text{LS0081}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0081}(1)(2)} N_{\text{LS0081}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0081}(930)(1)} N_{\text{LS0081}(930)(1)} \left( \hat{\mu}_{(930)(1)}, \hat{\sigma}_{(930)(1)}^2 \right) + \hat{w}_{\text{LS0081}(930)(2)} N_{\text{LS0081}(930)(2)} \left( \hat{\mu}_{(930)(2)}, \hat{\sigma}_{(930)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{LS0752}) &= \left[ \hat{w}_{\text{LS0752}(1)(1)} N_{\text{LS0752}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0752}(1)(2)} N_{\text{LS0752}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0752}(929)(1)} N_{\text{LS0752}(929)(1)} \left( \hat{\mu}_{(929)(1)}, \hat{\sigma}_{(929)(1)}^2 \right) + \hat{w}_{\text{LS0752}(929)(2)} N_{\text{LS0752}(929)(2)} \left( \hat{\mu}_{(929)(2)}, \hat{\sigma}_{(929)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{LS0688}) &= \left[ \hat{w}_{\text{LS0688}(1)(1)} N_{\text{LS0688}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{LS0688}(1)(2)} N_{\text{LS0688}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{LS0688}(934)(1)} N_{\text{LS0688}(934)(1)} \left( \hat{\mu}_{(934)(1)}, \hat{\sigma}_{(934)(1)}^2 \right) + \hat{w}_{\text{LS0688}(934)(2)} N_{\text{LS0688}(934)(2)} \left( \hat{\mu}_{(934)(2)}, \hat{\sigma}_{(934)(2)}^2 \right) \right] \\
 f_{\text{BMMA}}(\text{U273}) &= \left[ \hat{w}_{\text{U273}(1)(1)} N_{\text{U273}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{U273}(1)(2)} N_{\text{U273}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots + \\
 &\quad \left[ \hat{w}_{\text{U273}(936)(1)} N_{\text{U273}(936)(1)} \left( \hat{\mu}_{(936)(1)}, \hat{\sigma}_{(936)(1)}^2 \right) + \hat{w}_{\text{U273}(936)(2)} N_{\text{U273}(936)(2)} \left( \hat{\mu}_{(936)(2)}, \hat{\sigma}_{(936)(2)}^2 \right) \right]
 \end{aligned}$$

## Lampiran 9. (Lanjutan)

$$f_{\text{BMMA}}(\text{U325}) = \left[ \hat{w}_{\text{U325}(1)(1)} N_{\text{U325}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{U325}(1)(2)} N_{\text{U325}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots +$$

$$\left[ \hat{w}_{\text{U325}(928)(1)} N_{\text{U325}(928)(1)} \left( \hat{\mu}_{(928)(1)}, \hat{\sigma}_{(928)(1)}^2 \right) + \hat{w}_{\text{U325}(928)(2)} N_{\text{U325}(928)(2)} \left( \hat{\mu}_{(928)(2)}, \hat{\sigma}_{(928)(2)}^2 \right) \right]$$

$$f_{\text{BMMA}}(\text{U385}) = \left[ \hat{w}_{\text{U385}(1)(1)} N_{\text{U385}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{U385}(1)(2)} N_{\text{U385}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots +$$

$$\left[ \hat{w}_{\text{U385}(943)(1)} N_{\text{U385}(943)(1)} \left( \hat{\mu}_{(943)(1)}, \hat{\sigma}_{(943)(1)}^2 \right) + \hat{w}_{\text{U385}(943)(2)} N_{\text{U385}(943)(2)} \left( \hat{\mu}_{(943)(2)}, \hat{\sigma}_{(943)(2)}^2 \right) \right]$$

$$f_{\text{BMMA}}(\text{U447}) = \left[ \hat{w}_{\text{U447}(1)(1)} N_{\text{U447}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{U447}(1)(2)} N_{\text{U447}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots +$$

$$\left[ \hat{w}_{\text{U447}(922)(1)} N_{\text{U447}(922)(1)} \left( \hat{\mu}_{(922)(1)}, \hat{\sigma}_{(922)(1)}^2 \right) + \hat{w}_{\text{U447}(922)(2)} N_{\text{U447}(922)(2)} \left( \hat{\mu}_{(922)(2)}, \hat{\sigma}_{(922)(2)}^2 \right) \right]$$

$$f_{\text{BMMA}}(\text{CA0426}) = \left[ \hat{w}_{\text{CA0426}(1)(1)} N_{\text{CA0426}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{CA0426}(1)(2)} N_{\text{CA0426}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots +$$

$$\left[ \hat{w}_{\text{CA0426}(906)(1)} N_{\text{CA0426}(906)(1)} \left( \hat{\mu}_{(906)(1)}, \hat{\sigma}_{(906)(1)}^2 \right) + \hat{w}_{\text{CA0426}(906)(2)} N_{\text{CA0426}(906)(2)} \left( \hat{\mu}_{(906)(2)}, \hat{\sigma}_{(906)(2)}^2 \right) \right]$$

$$f_{\text{BMMA}}(\text{CA1120}) = \left[ \hat{w}_{\text{CA1120}(1)(1)} N_{\text{CA1120}(1)(1)} \left( \hat{\mu}_{(1)(1)}, \hat{\sigma}_{(1)(1)}^2 \right) + \hat{w}_{\text{CA1120}(1)(2)} N_{\text{CA1120}(1)(2)} \left( \hat{\mu}_{(1)(2)}, \hat{\sigma}_{(1)(2)}^2 \right) \right] + \dots +$$

$$\left[ \hat{w}_{\text{CA1120}(907)(1)} N_{\text{CA1120}(907)(1)} \left( \hat{\mu}_{(907)(1)}, \hat{\sigma}_{(907)(1)}^2 \right) + \hat{w}_{\text{CA1120}(907)(2)} N_{\text{CA1120}(907)(2)} \left( \hat{\mu}_{(907)(2)}, \hat{\sigma}_{(907)(2)}^2 \right) \right]$$

## 2. Model persamaan OLS untuk Model BUM untuk berbagai ID gen:

$$f_{\text{OLS}}(\text{LS0024}) = N_{\text{LS0024}} \left( \hat{\mu}_{\text{LS0024}}, \hat{\sigma}_{\text{LS0024}}^2 \right)$$

$$f_{\text{OLS}}(\text{LS0616}) = N_{\text{LS0616}} \left( \hat{\mu}_{\text{LS0616}}, \hat{\sigma}_{\text{LS0616}}^2 \right)$$

$$f_{\text{OLS}}(\text{LS0159}) = N_{\text{LS0159}} \left( \hat{\mu}_{\text{LS0159}}, \hat{\sigma}_{\text{LS0159}}^2 \right)$$

$$f_{\text{OLS}}(\text{LS0162}) = N_{\text{LS0162}} \left( \hat{\mu}_{\text{LS0162}}, \hat{\sigma}_{\text{LS0162}}^2 \right)$$

$$f_{\text{OLS}}(\text{LS0185}) = N_{\text{LS0185}} \left( \hat{\mu}_{\text{LS0185}}, \hat{\sigma}_{\text{LS0185}}^2 \right)$$

$$f_{\text{OLS}}(\text{LS0081}) = N_{\text{LS0081}} \left( \hat{\mu}_{\text{LS0081}}, \hat{\sigma}_{\text{LS0081}}^2 \right)$$

$$f_{\text{OLS}}(\text{LS0752}) = N_{\text{LS0752}} \left( \hat{\mu}_{\text{LS0752}}, \hat{\sigma}_{\text{LS0752}}^2 \right)$$

## Lampiran 9. (Lanjutan)

$$\begin{aligned}
f_{\text{OLS}}(\text{LS0688}) &= N_{\text{LS0688}} \left( \hat{\mu}_{\text{LS0688}}, \hat{\sigma}_{\text{LS0688}}^2 \right) \\
f_{\text{OLS}}(U070) &= N_{U070} \left( \hat{\mu}_{U070}, \hat{\sigma}_{U070}^2 \right) \\
f_{\text{OLS}}(U273) &= N_{U273} \left( \hat{\mu}_{U273}, \hat{\sigma}_{U273}^2 \right) \\
f_{\text{OLS}}(U001) &= N_{U001} \left( \hat{\mu}_{U001}, \hat{\sigma}_{U001}^2 \right) \\
f_{\text{OLS}}(U325) &= N_{U325} \left( \hat{\mu}_{U325}, \hat{\sigma}_{U325}^2 \right) \\
f_{\text{OLS}}(U385) &= N_{U385} \left( \hat{\mu}_{U385}, \hat{\sigma}_{U385}^2 \right) \\
f_{\text{OLS}}(U447) &= N_{U447} \left( \hat{\mu}_{U447}, \hat{\sigma}_{U447}^2 \right) \\
f_{\text{OLS}}(\text{CA0426}) &= N_{\text{CA0426}} \left( \hat{\mu}_{\text{CA0426}}, \hat{\sigma}_{\text{CA0426}}^2 \right) \\
f_{\text{OLS}}(\text{CA1120}) &= N_{\text{CA1120}} \left( \hat{\mu}_{\text{CA1120}}, \hat{\sigma}_{\text{CA1120}}^2 \right)
\end{aligned}$$



*Halaman ini sengaja dikosongkan*

## **DAFTAR RIWAYAT HIDUP**

### **I. IDENTITAS PRIBADI**



1. Nama : Ani Budi Astuti
2. NRP : 1311301002
3. TTL : Kediri, 9-2-1968
4. Agama : Islam
5. Instansi Asal : Universitas  
Brawijaya Jurusan  
Matematika/Program Studi Statistika
6. NIP : 196802091992032001
7. Rumah : Zona  
*Neighbourhood*  
NG/4 Sawojajar II  
Malang Jawa  
Timur
8. Email : ani\_budi@ub.ac.id
9. No. HP : 081334677001

## II. RIWAYAT PENDIDIKAN

No	Jenjang	Tempat Pendidikan	Bidang Studi/Keahlian	Tahun Mulai	Tahun Lulus
1	S2	IPB Bogor	Statistika	1994	1997
2	S1	Universitas Brawijaya Malang	Matematika	1987	1991
3	SMA	SMAN II Kediri	Biologi (A2)	1984	1987
4	SMP	SMPN Ngadiluwih Kediri	-	1981	1984
5	SD	SDN Ngadiluwih Kediri	-	1975	1981

## III. RIWAYAT PUBLIKASI

Publikasi yang telah dilakukan dalam bentuk jurnal dan seminar selama masa studi:

### Publikasi di Jurnal dan Proceeding Internasional:

1. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2014a). Model Components Selection in Bayesian Model Averaging Using Occam's Window for Microarray Data. *Natural-A Journal of Scientific Modeling and Computation*, 2 (1): 67-74. ISSN: 2303-0135.
2. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2014b). Kolmogorov-Smirnov and Continuous Ranked Probability Score Validation on the Bayesian Model Averaging for Microarray Data. *Applied Mathematical Sciences*, 146 (8): 7277-7287. DOI: <http://dx.doi.org/10.12988/ams.2014.49760>.

3. Astuti, A. B. (2014c). Cox Regression Model for Lifetime of a Newborn Baby (Neonates) in Dr. Saiful Anwar Malang Hospital. *Global Journal of Pure and Applied Mathematics*, 10 (4): 591-600. ISSN 0973-1768.
4. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015b). Occam's Window Selection in Bayesian Model Averaging Modeling for Gene Expression Data from *Chickpea* Plant. *International Journal of Applied Mathematics and Statistics*, 53 (4): 160-165. ISSN 0973-1377 (Print), ISSN 0973-7545 (Online).
5. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015c). An Algorithm for Determining the Number of Mixture Components on the Bayesian Mixture Model Averaging for Microarray Data. *Journal of Mathematics and Statistics*, 11 (2): 45-51. DOI: 10.3844/jmssp.2015.45.51.
6. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016a). Blood Sugar Levels of Diabetes Mellitus Patients Modeling with Bayesian Mixture Model Averaging. *Global Journal of Pure and Applied Mathematics*, 12 (4): 3143–3158. ISSN 0973-1768.
7. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016c). Bayesian Mixture Model Averaging (BMMA) for Identifying the Differences Genes Expression of *Chickpea* (*Cicer Arietinum*) Plant Tissue. *Communications in Statistics-Theory and Methods*, Publish Online Accepted Manuscript Version. DOI: <http://dx.doi.org/10.1080/03610926.2016.1239112>.
8. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016d). Bayesian Mixture Modeling for Blood Sugar Levels of Diabetes Mellitus Patients (Case Study in RSUD Saiful Anwar Malang Indonesia). *Accepted publishon Proceeding of IEEEExplore 2017*.

#### **Publikasi dan Diseminasi di Seminar Internasional dan Nasional:**

1. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2012). Theoretical Studies of Bayesian Mixture Model Averaging Methods for Microarray Data. *Proceeding of 2<sup>nd</sup> Regional Conference on Applied and Engineering Mathematics (RCAEM-II)*. Penang, Malaysia. ISBN 978-967-11189-00.
2. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2013). Highest Posterior Density for Identifying Differences in Gene Expression Microarray Experiments. *Proceeding of the 3<sup>rd</sup> Annual Basic Science International Conference (BaSIC 2013)*. Volume 3. UB, Malang. ISSN 2338-0136.
3. Astuti, A. B., Iriawan N., Irhamah, and Kuswanto H. (2014). Model Components Selection in Bayesian Model Averaging using Occam's Window for Microarray Data. The 4<sup>th</sup> Annual Basic Science International Conference 2014 (BaSic 2014) in conjunction with The 5<sup>th</sup> International Conference on Global Resource Conservation 2014. FMIPA UB, Malang. ISSN 2338-0136.

4. Astuti, A. B., Iriawan N., Irhamah, and Kuswanto H. (2014). Occam's Window Selection in Bayesian Model Averaging Modeling for Gene Expression Data from *Chickpea* Plant. 2014 International Conference on Statistics and Mathematics (ICSM 2014). Department of Statistics, FMIPA ITS, Surabaya.
5. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015). Bayesian Model Averaging dengan Metode Seleksi Komponen Model Occam's Window pada Data Diabetes Mellitus. Seminar Nasional SEMIRATA Bidang MIPA 2015. Universitas Tanjungpura, Pontianak.
6. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015). Bayesian Model Averaging Modeling for Blood Sugar Levels of Diabetes Mellitus Patients in RSUD Saiful Anwar Malang. *International Symposium on Biomathematics (Symomath)* 2015. ITB, Bandung.
7. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2015). Blood Sugar Levels of Diabetes Mellitus Patients Modeling with Bayesian Mixture Model Averaging. *The First International Conference on Statistical Methods in Engineering, Science, Economy, and Education*. UIN, Yogyakarta.
8. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016). Bayesian Mixture Modeling for Blood Sugar Levels of Diabetes Mellitus Patients (Case Study in RSUD Saiful Anwar Malang Indonesia). *The Asian Mathematical Conference (AMC)* 2016. ITB, Bandung.
9. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016). Goodness of Fits Model with Kolmogorov-Smirnov on the Bayesian Mixture Model Averaging (Case Study on Diabetes Mellitus Patients at RSUD Saiful Anwar Malang). *International Symposium on Biomathematics (Symomath)* 2016. Universitas Hassanudin, Makassar.
10. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2016e). Seleksi Model Occam's Window dalam Pemodelan Bayesian Model Averaging (Studi Kasus Diabetes Mellitus di RSUD Saiful Anwar Malang). Konferensi Nasional Matematika (KNM) ke-18 2016. Universitas Riau, Pekanbaru.
11. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2017). Algoritma *Reversible Jump Markov Chain Monte Carlo* dalam Pemodelan Bayesian untuk Mengidentifikasi Jumlah Komponen *Mixture* pada Data Diabetes Mellitus di Indonesia. Konferensi Nasional Matematika dan Aplikasinya (SEMNAS MANTAP I) 2017. Universitas Lambung Mangkurat, Banjarbaru, Kalimantan Selatan.

12. Astuti, A. B., Iriawan N., Irhamah, dan Kuswanto H. (2017). Development of Reversible Jump Markov Chain Monte Carlo Algorithm in the Bayesian Mixture Modeling for Microarray Data in Indonesia. *International Conference & Workshop on Mathematical Analysis and Its Applications (ICWOMAA) 2017*. UB, Malang.

#### **IV. RIWAYAT PENELITIAN**

1. Penelitian Unggulan Perguruan Tinggi (PUPT-BOPTN) tahun 2015. Pemodelan Kadar Gula Darah Penderita Penyakit *Diabetes Mellitus* melalui Pengembangan Algoritma Bayesian Mixture Model Averaging. Sebagai Ketua.
2. Penelitian Unggulan Perguruan Tinggi (PUPT-BOPTN) tahun 2016. Pemodelan Kadar Gula Darah Penderita Penyakit *Diabetes Mellitus* melalui Pengembangan Algoritma Bayesian Mixture Model Averaging. Sebagai Ketua.

#### **V. RIWAYAT PENGABDIAN KEPADA MASYARAKAT**

1. Sebagai juara II dalam kegiatan kompetisi pembuatan Proposal Pengabdian Kepada Masyarakat pada kegiatan Workshop dan Pendampingan Penyusunan Proposal Pengabdian Kepada Masyarakat bagi Dosen Angkatan II yang diselenggarakan oleh Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Brawijaya pada tanggal 27-28 dan 30 Oktober 2014 di Malang. Sebagai Anggota.
2. IPTEKS bagi Masyarakat (IbM) tahun 2015. IbM Guru-Guru SMP dan MTs di Kabupaten Ponorogo dengan Pelatihan Perangkat Lunak Matematika sebagai Metode Pembelajaran. Sebagai Anggota.

#### **VI. LAIN-LAIN**

1. Mengikuti Spatial Statistics Workshop yang diadakan oleh Department of Statistics, FMIPA, IPB bekerja sama dengan Faculty of Spatial Sciences, University of Groningen, Netherlands pada tanggal 17-20 September 2012 di Bogor.
2. Mengikuti Workshop on Spatial and Spatial Time Modeling yang diadakan oleh Program Pascasarjana, FMIPA, UB bekerja sama dengan Faculty of Spatial Sciences, University of Groningen, Netherlands pada tanggal 17-18 Januari 2013 di Malang.
3. Sebagai Wakil Ketua Pelaksana Workshop Pengajaran Statistika yang diselenggarakan oleh Program Studi Statistika Jurusan Matematika

FMIPA UB bekerjasama dengan FPTSI, IPB, UGM dan STIS pada tanggal 12-14 April 2013 di Malang.

4. Mengikuti SAS Training (SAS Enterprise Guide 1: Querying and Reporting, SAS Programming 1: Essentials, SAS Enterprise Guide: ANOVA, Regression, and Logistic Regression) yang diselenggarakan oleh Program Studi Statistika FMIPA UB bekerjasama dengan SAS pada tanggal 16-20 Desember 2013 di Malang.
5. Mengikuti Konsinyering Penulisan dan Publikasi Artikel pada Jurnal Internasional bagi Dosen yang diselenggarakan oleh PPKID Universitas Brawijaya pada tanggal 9-10 Juni 2014 di Malang.
6. Workshop dan Pendampingan Penyusunan Proposal Pengabdian Kepada Masyarakat bagi Dosen Angkatan II yang diselenggarakan oleh Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Brawijaya pada tanggal 27-28 dan 30 Oktober 2014 di Malang.
7. Mengikuti Pelatihan Penulisan Artikel untuk Jurnal Internasional dan Proposal Penelitian yang diselenggarakan oleh the Indonesian Mathematical Society (IndoMS) Wilayah Jawa Timur dan CRDF Global pada tanggal 15 Nopember 2014 di Surabaya.
8. Mengikuti International Workshop on Bayesian Modelling yang diadakan oleh Department of Statistics, FMIPA, ITS bekerja sama dengan Queensland University of Technology Australia and Flinders University Australia pada tanggal 24-26 Nopember 2014 di Surabaya.
9. Sebagai panitia penyelenggara dan sekaligus peserta International Workshop on Bayesian Modeling, Theory and Its Application yang diselenggarakan di Program Studi Statistika, Jurusan Matematika, FMIPA UB Malang pada tanggal 29 February 2016 dan 01 March 2016 dengan promotor sebagai Instruktur kegiatan tersebut. Kegiatan tersebut sebagai rangkaian kegiatan BaSIC 2016 di Universitas Brawijaya Malang.
10. Sebagai panitia kegiatan Seminar Internasional BaSIC UB tanggal 2-3 Maret 2016 dimana Jurusan Matematika FMIPA UB Malang sebagai ketua pelaksana kegiatan.
11. Mendaftar sebagai peserta Kegiatan Lokakarya/Konsinyering Penulisan dan Publikasi Artikel pada Jurnal Internasional tanggal 25-26 Mei 2016 di Universitas Brawijaya (UB) Malang dengan mengirimkan draft paper yang akan dipublikasi ke salah satu jurnal internasional yang akan dituju.
12. Mendapatkan Travel Grants dari Japan Mathematical Society sebesar 300 \$ US pada kegiatan AMC 2016 di Bali yang diselenggarakan pada tanggal 25-29 July 2016.

13. Mengikuti kompetisi Insentif Jurnal Internasional 2016 yang diselenggarakan oleh DIKTI, judul paper yang diajukan “An Algorithm for Determining the Number of Mixture Components on the Bayesian Mixture Model Averaging for Microarray Data”, yang telah terbit di jurnal internasional index scopus JMS (*Journal of Mathematics and Statistics*) Vol. 11, Issue No. 2, Year 2015, Halaman 45-51. DOI: 10.3844/jmssp.2015.45.51.
14. Menjadi reviewer di tiga Jurnal Internasional, yaitu MASA (*Model Assisted Statistics and Applications*-IOS Press Publisher-Netherlands), AJAS (*American Journal of Applied Sciences*- Science Publications Publisher-US) dan JMS (*Journal of Mathematics and Statistics*-Science Publications Publisher-US).
15. Mengikuti kegiatan *Joint Posgraduate Seminar*, antara PSU-Thailand dengan ITS sebagai *Contributed Speaker* di Ruang Sidang Rektorat Lantai 1 ITS dan kegiatan *Workshop Forecasting* di Laboratorium Komputer Lantai 3, Jurusan Statistika ITS, serta peserta Seminar Internasional ICTAS di Hotel Majapahit pada tanggal 17-20 Oktober 2016.

Surabaya, 28 Juli 2017

Ani Budi Astuti  
NRP. 1311301002